Name: _____

GCE Stats section 1

Numerical measures, graphs and diagrams

practice questions from GCE Maths

Date:

Question parts in red cover content in the GCE Statistics specification but outside this section of the scheme of work

Time:

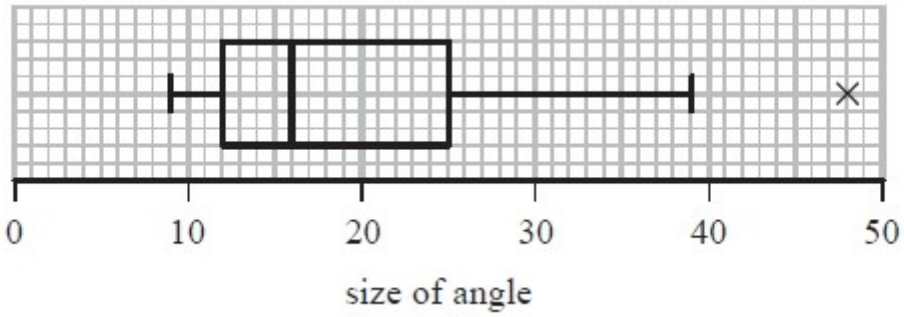
Total marks available:

Total marks achieved: _____

Questions

Q1.

Each of 60 students was asked to draw a 20° angle without using a protractor. The size of
each angle drawn was measured. The results are summarised in the box plot below.



(a) Find the range for these data.

(1)

(b) Find the interquartile range for these data.

(1)

The students were then asked to draw a 70° angle.
The results are summarised in the table below.

Angle, a, (degrees)	Number of students
$55 \leqslant a < 60$	6
$60 \leqslant a < 65$	15
$65 \leqslant a < 70$	13
$70 \leqslant a < 75$	11
$75 \leqslant a < 80$	8
$80 \leqslant a < 85$	7

(c) Use linear interpolation to estimate the size of the median angle drawn. Give your
answer to 1 decimal place.

(2)

(d) Show that the lower quartile is 63°

(2)

For these data, the upper quartile is 75°, the minimum is 55° and the maximum is 84°

An outlier is an observation that falls either
more than 1.5 × (interquartile range) above the upper quartile or
more than 1.5 × (interquartile range) below the lower quartile.

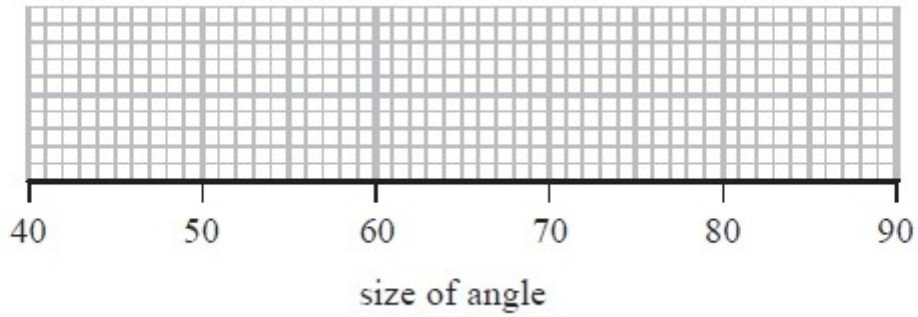
(e) (i) Show that there are no outliers for these data.

(ii) Draw a box plot for these data on the grid on page 3.

(5)

(f) State which angle the students were more accurate at drawing. Give reasons for your answer.

(3)



(Total for question = 14 marks)

Q2.

A random sample of 35 homeowners was taken from each of the villages Greenslax and Penville and their ages were recorded. The results are summarised in the back-to-back stem and leaf diagram below.

Totals	Greenslax									Penville											Totals
(2)							8	7	2	5	5	6	7	8	8	9					(7)
(3)						9	8	7	3	1	1	1	2	3	4	4	5	6	7	9	(11)
(4)					4	4	4	0	4	0	1	2	4	7							(5)
(5)				6	6	5	2	2	5	0	0	5	5	5							(5)
(7)		8	6	5	4	2	1	1	6	2	5	6	6								(4)
(8)	8	6	6	6	4	3	1	1	7	0	5										(2)
(5)				9	8	4	3	2	8												(0)
(1)								4	9	9											(1)

Key: 7 | 3 | 1 means 37 years for Greenslax and 31 years for Penville

Some of the quartiles for these two distributions are given in the table below.

	Greenslax	Penville
Lower quartile, Q_1	a	31
Median, Q_2	64	39
Upper quartile, Q_3	b	55

(a) Find the value of a and the value of b.

(2)

An outlier is a value that falls either

more than $1.5 \times (Q_3 - Q_1)$ above Q_3

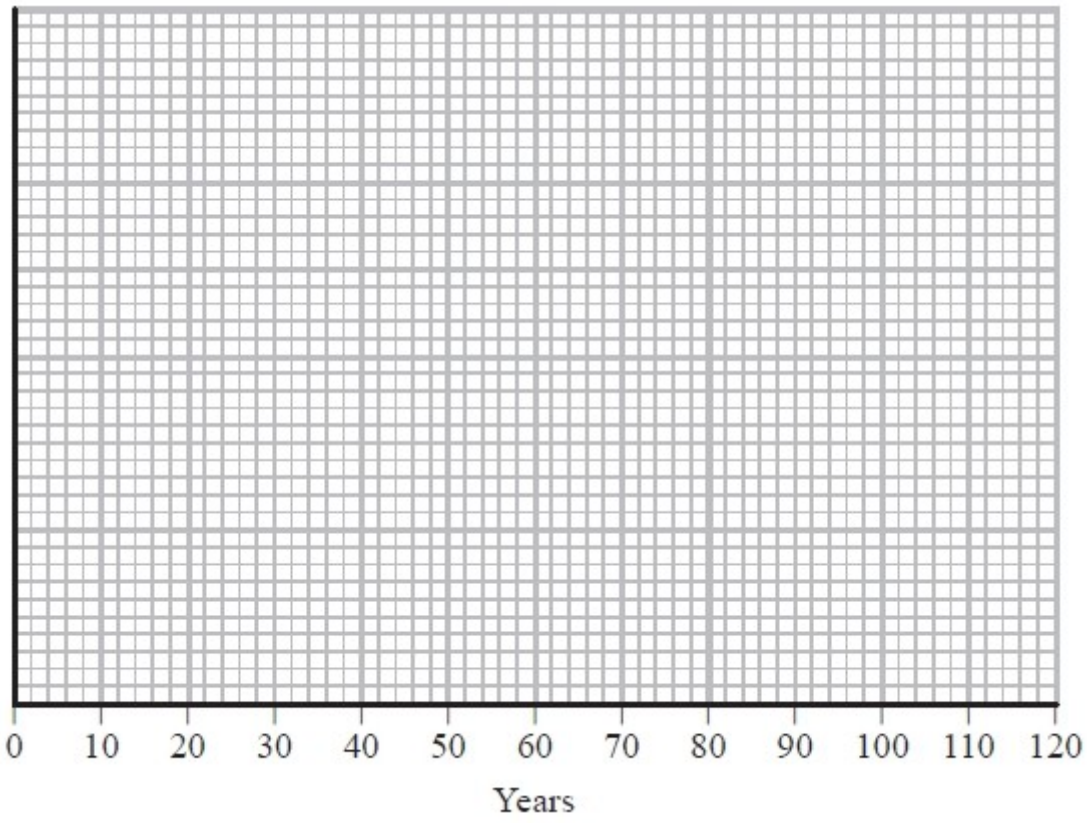
or more than $1.5 \times (Q_3 - Q_1)$ below Q_1

(b) On the graph paper opposite draw a box plot to represent the data from Penville.
Show clearly any outliers.

(4)

(c) State the skewness of each distribution. Justify your answers.

(3)



(Total 9 marks)

Q3.

The histogram in Figure 1 shows the time, to the nearest minute, that a random sample of 100 motorists were delayed by roadworks on a stretch of motorway.

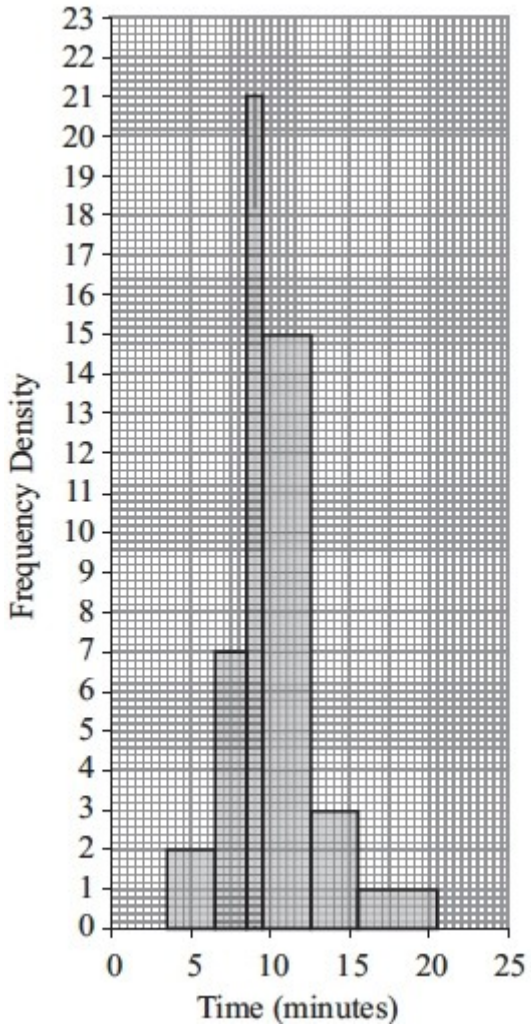


Figure 1

(a) Complete the table.

Delay (minutes)	Number of motorists
4 – 6	6
7 – 8	
9	21
10 – 12	45
13 – 15	9
16 – 20	

(2)

(b) Estimate the number of motorists who were delayed between 8.5 and 13.5 minutes by the roadworks.

(2)

(Total 4 marks)

Q4.

An estate agent is studying the cost of office space in London. He takes a random sample of 90 offices and calculates the cost, £x per square foot. His results are given in the table below.

Cost (£x)	Frequency (f)	Midpoint (£y)
$20 \leqslant x < 40$	12	30
$40 \leqslant x < 45$	13	42.5
$45 \leqslant x < 50$	25	47.5
$50 \leqslant x < 60$	32	55
$60 \leqslant x < 80$	8	70

(You may use $\sum f y^2 = 226\,687.5$)

A histogram is drawn for these data and the bar representing $50 \leqslant x < 60$ is 2cm wide and 8cm high.

(a) Calculate the width and height of the bar representing $20 \leqslant x < 40$

(3)

(b) Use linear interpolation to estimate the median cost.

(2)

(c) Estimate the mean cost of office space for these data.

(2)

(d) Estimate the standard deviation for these data.

(2)

(e) Describe, giving a reason, the skewness.

(1)

Rika suggests that the cost of office space in London can be modelled by a normal distribution
with mean £50 and standard deviation £10

(f) With reference to your answer to part (e), comment on Rika's suggestion.

(1)

(g) Use Rika's model to estimate the 80th percentile of the cost of office space in London.

(3)

(Total for question = 14 marks)

Q5.

The marks of a group of female students in a statistics test are summarised in Figure 1

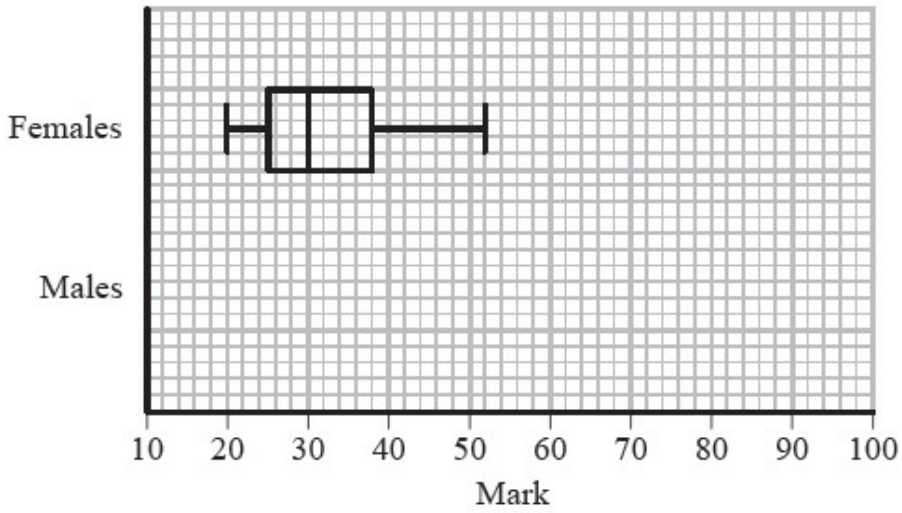


Figure 1

(a) Write down the mark which is exceeded by 75% of the female students.

(1)

The marks of a group of male students in the same statistics test are summarised by the stem and leaf diagram below.

Mark	(2|6 means 26)	Totals
1	4	(1)
2	6	(1)
3	4 4 7	(3)
4	0 6 6 7 7 8	(6)
5	0 0 1 1 1 3 6 7 7	(9)
6	2 2 3 3 3 8	(6)
7	0 0 8	(3)
8	5	(1)
9	0	(1)

(b) Find the median and interquartile range of the marks of the male students.

(3)

An outlier is a mark that is

either more than 1.5 × interquartile range above the upper quartile

or more than 1.5 × interquartile range below the lower quartile.

(c) In the space provided on Figure 1 draw a box plot to represent the marks of the male students, indicating clearly any outliers.

(5)

(d) Compare and contrast the marks of the male and the female students.

(2)

(Total 11 marks)

Q6.

Keith records the amount of rainfall, in mm, at his school, each day for a week. The results are given below.

2.8 5.6 2.3 9.4 0.0 0.5 1.8

Jenny then records the amount of rainfall, x mm, at the school each day for the following 21 days.
The results for the 21 days are summarised below.

$$\sum x = 84.6$$

(a) Calculate the mean amount of rainfall during the whole 28 days.

(2)

Keith realises that he has transposed two of his figures. The number 9.4 should have been 4.9 and
the number 0.5 should have been 5.0
Keith corrects these figures.

(b) State, giving your reason, the effect this will have on the mean.

(2)

(Total 4 marks)

Q7.

Before going on holiday to *Seapron*, Tania records the weekly rainfall (x mm) at *Seapron* for 8 weeks during the summer. Her results are summarised as

$$\sum x = 86.8 \qquad \sum x^2 = 985.88$$

(a) Find the standard deviation, σ_x, for these data.

(3)

Tania also records the number of hours of sunshine (y hours) per week at *Seapron* for these 8 weeks and obtains the following

$$\bar{y} = 58 \qquad \sigma_y = 9.461 \text{ (correct to 4 significant figures)} \qquad \sum xy = 4900.5$$

(b) Show that $S_{yy} = 716$ (correct to 3 significant figures)

(1)

(c) Find S_{xy}

(2)

(d) Calculate the product moment correlation coefficient, r, for these data.

(2)

During Tania's week-long holiday at *Seapron* there are 14 mm of rain and 70 hours of sunshine.

(e) State, giving a reason, what the effect of adding this information to the above data would be on the value of the product moment correlation coefficient.

(2)

(Total for question = 10 marks)

Q8.

An agriculturalist is studying the yields, y kg, from tomato plants. The data from a random sample of
70 tomato plants are summarised below.

Yield (y kg)	**Frequency** (f)	**Yield midpoint** (x kg)
$0 \leqslant y < 5$	16	2.5
$5 \leqslant y < 10$	24	7.5
$10 \leqslant y < 15$	14	12.5
$15 \leqslant y < 25$	12	20
$25 \leqslant y < 35$	4	30

(You may use $\sum fx = 755$ and $\sum fx^2 = 12037.5$)

A histogram has been drawn to represent these data.

The bar representing the yield $5 \leqslant y < 10$ has a width of 1.5 cm and a height of 8 cm.

(a) Calculate the width and the height of the bar representing the yield $15 \leqslant y < 25$

(3)

(b) Use linear interpolation to estimate the median yield of the tomato plants.

(2)

(c) Estimate the mean and the standard deviation of the yields of the tomato plants.

(4)

(d) Describe, giving a reason, the skewness of the data.

(2)

(e) Estimate the number of tomato plants in the sample that have a yield of more than 1 standard deviation above the mean.

(2)

(Total 13 marks)

Q9.

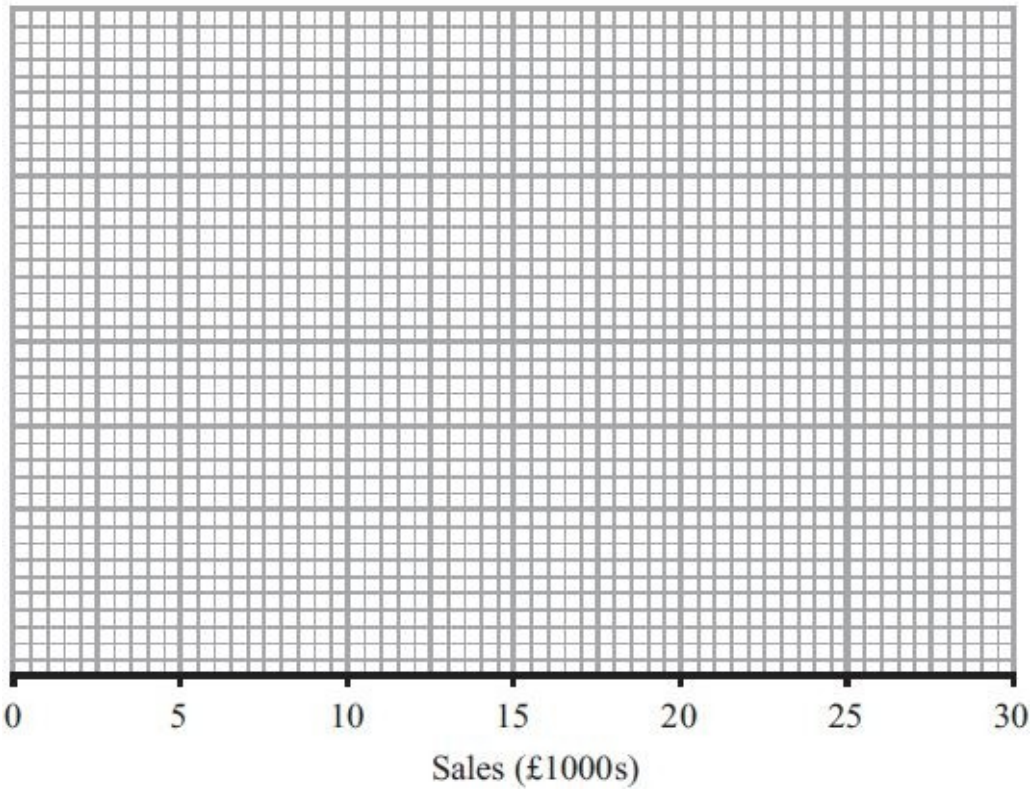
Over a long period of time a small company recorded the amount it received in sales per month.
The results are summarised below.

	Amount received in sales (£1000s)
Two lowest values	3, 4
Lower quartile	7
Median	12
Upper quartile	14
Two highest values	20, 25

An outlier is an observation that falls
either 1.5 × interquartile range above the upper quartile
or 1.5 × interquartile range below the lower quartile.

(a) On the graph paper below, draw a box plot to represent these data, indicating clearly any outliers.

(5)



(b) State the skewness of the distribution of the amount of sales received. Justify your answer.

(2)

(c) The company claims that for 75% of the months, the amount received per month is greater

than £10 000. Comment on this claim, giving a reason for your answer.

(2)

(Total 9 marks)

Q10.

The following table summarises the times, t minutes to the nearest minute, recorded for a group of students to complete an exam.

Time (minutes) t	11 – 20	21 – 25	26 – 30	31 – 35	36 – 45	46 – 60
Number of students f	62	88	16	13	11	10

[You may use $\sum ft^2 = 134281.25$]

(a) Estimate the mean and standard deviation of these data.

(5)

(b) Use linear interpolation to estimate the value of the median.

(2)

(c) Show that the estimated value of the lower quartile is 18.6 to 3 significant figures.

(1)

(d) Estimate the interquartile range of this distribution.

(2)

(e) Give a reason why the mean and standard deviation are not the most appropriate summary statistics to use with these data.

(1)

The person timing the exam made an error and each student actually took 5 minutes less than the times recorded above. The table below summarises the actual times.

Time (minutes) t	6 – 15	16 – 20	21 – 25	26 – 30	31 – 40	41 – 55
Number of students f	62	88	16	13	11	10

(f) Without further calculations, explain the effect this would have on each of the estimates found in
parts (a), (b), (c) and (d).

(3)

(Total 14 marks)

Q11.

The marks, *x*, of 45 students randomly selected from those students who sat a mathematics examination are shown in the stem and leaf diagram below.

Mark		Totals
3	6 9 9	(3)
4	0 1 2 2 3 4	(6)
4	5 6 6 6 8	(5)
5	0 2 3 3 4 4	(6)
5	5 5 6 7 7 9	(6)
6	0 0 0 0 1 3 4 4 4	(9)
6	5 5 6 7 8 9	(6)
7	1 2 3 3	(4)

Key	(3|6 means 36)

(a) Write down the modal mark of these students.

(1)

(b) Find the values of the lower quartile, the median and the upper quartile.

(3)

For these students $\sum x$ = 2497 and $\sum x^2$ = 143 369

(c) Find the mean and the standard deviation of the marks of these students.

(3)

(d) Describe the skewness of the marks of these students, giving a reason for your answer.

(2)

The mean and standard deviation of the marks of all the students who sat the examination were 55
and 10 respectively. The examiners decided that the total mark of each student should be scaled by subtracting 5 marks and then reducing the mark by a further 10%.

(e) Find the mean and standard deviation of the scaled marks of all the students.

(4)

(Total 13 marks)

Q12.

A midwife records the weights, in kg, of a sample of 50 babies born at a hospital. Her results are given in the table below.

Weight (w kg)	Frequency (f)	Weight midpoint (x)
$0 \leqslant w < 2$	1	1
$2 \leqslant w < 3$	8	2.5
$3 \leqslant w < 3.5$	17	3.25
$3.5 \leqslant w < 4$	17	3.75
$4 \leqslant w < 5$	7	4.5

[You may use $\sum fx^2 = 611.375$]

A histogram has been drawn to represent these data.

The bar representing the weight $2 \leqslant w < 3$ has a width of 1 cm and a height of 4 cm.

(a) Calculate the width and height of the bar representing a weight of $3 \leqslant w < 3.5$

(3)

(b) Use linear interpolation to estimate the median weight of these babies.

(2)

(c) (i) Show that an estimate of the mean weight of these babies is 3.43 kg.

(ii) Find an estimate of the standard deviation of the weights of these babies.

(3)

Shyam decides to model the weights of babies born at the hospital, by the random variable W,
where $W \sim N(3.43, 0.65^2)$

(d) Find $P(W < 3)$

(3)

(e) With reference to your answers to (b), (c)(i) and (d) comment on Shyam's decision.

(3)

A newborn baby weighing 3.43 kg is born at the hospital.

(f) Without carrying out any further calculations, state, giving a reason, what effect the addition of this newborn baby to the sample would have on your estimate of the

(i) mean,
(ii) standard deviation.

(3)

(Total for question = 17 marks)

Q13.

The table shows the time, to the nearest minute, spent waiting for a taxi by each of 80 people one
Sunday afternoon.

Waiting time (in minutes)	Frequency
2–4	15
5–6	9
7	6
8	24
9–10	14
11–15	12

(a) Write down the upper class boundary for the 2–4 minute interval.

(1)

A histogram is drawn to represent these data. The height of the tallest bar is 6 cm.

(b) Calculate the height of the second tallest bar.

(3)

(c) Estimate the number of people with a waiting time between 3.5 minutes and 7 minutes.

(2)

(d) Use linear interpolation to estimate the median, the lower quartile and the upper quartile of the
waiting times.

(4)

(e) Describe the skewness of these data, giving a reason for your answer.

(2)

(Total 12 marks)

Q14.

A survey of 100 households gave the following results for weekly income £y.

Income y (£)	Mid-point	Frequency f
$0 \leqslant y < 200$	100	12
$200 \leqslant y < 240$	220	28
$240 \leqslant y < 320$	280	22
$320 \leqslant y < 400$	360	18
$400 \leqslant y < 600$	500	12
$600 \leqslant y < 800$	700	8

(You may use $\Sigma fy^2 = 12\,452\,800$)

A histogram was drawn and the class $200 \leqslant y < 240$ was represented by a rectangle of width 2 cm
and height 7 cm.

(a) Calculate the width and the height of the rectangle representing the class $320 \leqslant y < 400$

(3)

(b) Use linear interpolation to estimate the median weekly income to the nearest pound.

(2)

(c) Estimate the mean and the standard deviation of the weekly income for these data.

(4)

One measure of skewness is $\frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$.

(d) Use this measure to calculate the skewness for these data and describe its value.

(2)

Katie suggests using the random variable X which has a normal distribution with mean 320 and
standard deviation 150 to model the weekly income for these data.

(e) Find $P(240 < X < 400)$.

(2)

(f) With reference to your calculations in parts (d) and (e) and the data in the table, comment on Katie's suggestion.

(2)

(Total 15 marks)

Q15.

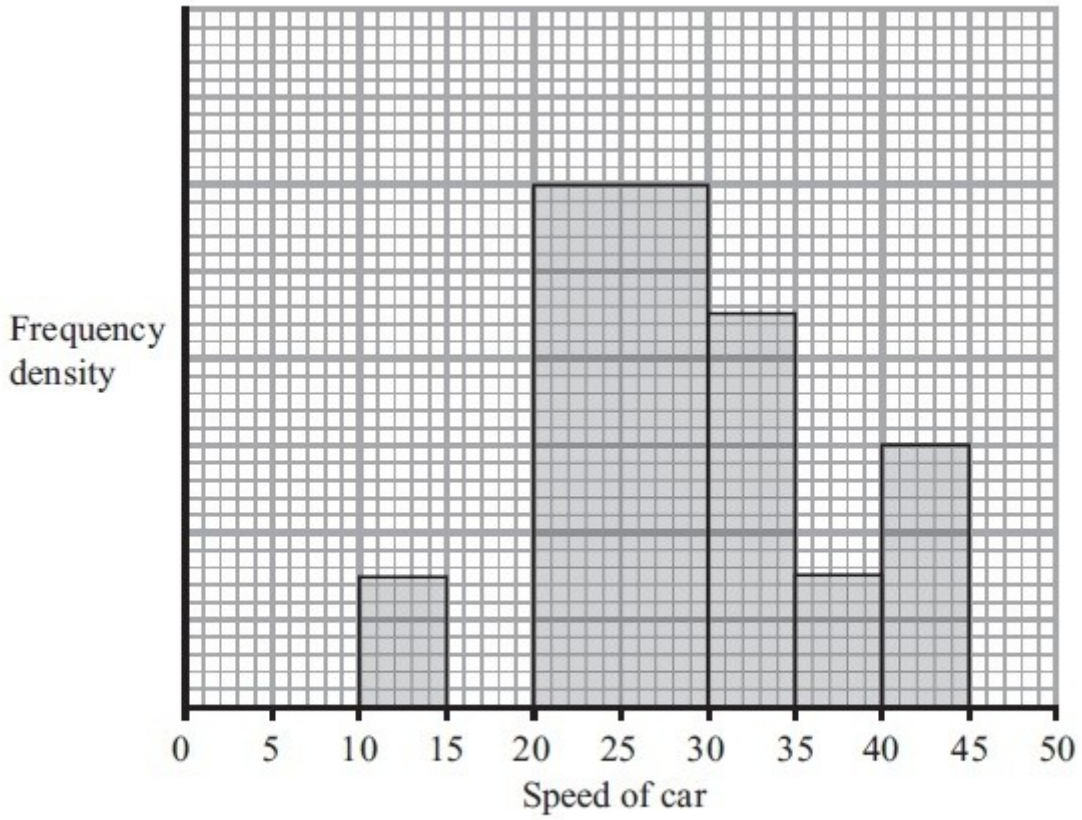


Figure 2

A policeman records the speed of the traffic on a busy road with a 30 mph speed limit. He records the speeds of a sample of 450 cars. The histogram in Figure 2 represents the results.

(a) Calculate the number of cars that were exceeding the speed limit by at least 5 mph in the sample.

(4)

(b) Estimate the value of the mean speed of the cars in the sample.

(3)

(c) Estimate, to 1 decimal place, the value of the median speed of the cars in the sample.

(2)

(d) Comment on the shape of the distribution. Give a reason for your answer.

(2)

(e) State, with a reason, whether the estimate of the mean or the median is a better representation
of the average speed of the traffic on the road.

(2)

(Total 13 marks)

Q16.

A class of students had a sudoku competition. The time taken for each student to complete the sudoku was recorded to the nearest minute and the results are summarised in the table below.

Time	Mid-point, x	Frequency, f
2 - 8	5	2
9 - 12		7
13 - 15	14	5
16 - 18	17	8
19 - 22	20.5	4
23 - 30	26.5	4

(You may use $\Sigma fx^2 = 8603.75$)

(a) Write down the mid-point for the 9 - 12 interval.

(1)

(b) Use linear interpolation to estimate the median time taken by the students.

(2)

(c) Estimate the mean and standard deviation of the times taken by the students.

(5)

The teacher suggested that a normal distribution could be used to model the times taken by the students to complete the sudoku.

(d) Give a reason to support the use of a normal distribution in this case.

(1)

On another occasion the teacher calculated the quartiles for the times taken by the students to complete
a different sudoku and found

$$Q_1 = 8.5 \quad Q_2 = 13.0 \quad Q_3 = 21.0$$

(e) Describe, giving a reason, the skewness of the times on this occasion.

(2)

(Total 11 marks)

Q17.

On a randomly chosen day, each of the 32 students in a class recorded the time, t minutes to the nearest minute, they spent on their homework. The data for the class is summarised in the following table.

Time, t	Number of students
10 – 19	2
20 – 29	4
30 – 39	8
40 – 49	11
50 – 69	5
70 – 79	2

(a) Use interpolation to estimate the value of the median.

(2)

Given that

$$\sum t = 1414 \quad \text{and} \quad \sum t^2 = 69378$$

(b) find the mean and the standard deviation of the times spent by the students on their homework.

(3)

(c) Comment on the skewness of the distribution of the times spent by the students on their homework. Give a reason for your answer.

(2)

(Total 7 marks)

Q18.

A teacher selects a random sample of 56 students and records, to the nearest hour, the time spent watching television in a particular week.

Hours	1–10	11–20	21–25	26–30	31–40	41–59
Frequency	6	15	11	13	8	3
Mid-point	5.5	15.5		28		50

(a) Find the mid-points of the 21-25 hour and 31-40 hour groups.

(2)

A histogram was drawn to represent these data. The 11-20 group was represented by a bar of width 4 cm and height 6 cm.

(b) Find the width and height of the 26-30 group.

(3)

(c) Estimate the mean and standard deviation of the time spent watching television by these students.

(5)

(d) Use linear interpolation to estimate the median length of time spent watching television by
these students.

(2)

The teacher estimated the lower quartile and the upper quartile of the time spent watching television
to be 15.8 and 29.3 respectively.

(e) State, giving a reason, the skewness of these data.

(2)

(Total 14 marks)

Q19.

The times, in seconds, spent in a queue at a supermarket by 85 randomly selected customers,
are summarised in the table below.

Time (seconds)	Number of customers, f
$0 - 30$	2
$30 - 60$	10
$60 - 70$	17
$70 - 80$	25
$80 - 100$	25
$100 - 150$	6

A histogram was drawn to represent these data. The 30 – 60 group was represented by a bar of
width 1.5 cm and height 1 cm.

(a) Find the width and the height of the 70 – 80 group.

(3)

(b) Use linear interpolation to estimate the median of this distribution.

(2)

Given that x denotes the midpoint of each group in the table and

$$\sum fx = 6460 \qquad \sum fx^2 = 529\,400$$

(c) calculate an estimate for

(i) the mean,
(ii) the standard deviation,
for the above data.

(3)

One measure of skewness is given by

$$\text{coefficient of skewness} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

(d) Evaluate this coefficient and comment on the skewness of these data.

(3)

(Total 11 marks)

Q20.

The distances travelled to work, D km, by the employees at a large company are normally distributed

with $D \sim N(30, 8^2)$.

(a) Find the probability that a randomly selected employee has a journey to work of more than 20 km.

(3)

(b) Find the upper quartile, Q_3, of D.

(3)

(c) Write down the lower quartile, Q_1, of D.

(1)

An outlier is defined as any value of D such that $D < h$ or $D > k$ where

$$h = Q_1 - 1.5 \times (Q_3 - Q_1) \quad \text{and} \quad k = Q_3 + 1.5 \times (Q_3 - Q_1)$$

(d) Find the value of h and the value of k.

(2)

An employee is selected at random.

(e) Find the probability that the distance travelled to work by this employee is an outlier.

(3)

(Total 12 marks)

Q21.

A student is investigating the numbers of cherries in a *Rays* fruit cake. A random sample of *Rays* fruit cakes is taken and the results are shown in the table below.

Number of cherries	0	1	2	3	4	5	$\geqslant 6$
Frequency	24	37	21	12	4	2	0

(a) Calculate the mean and the variance of these data.

(3)

(b) Explain why the results in part (a) suggest that a Poisson distribution may be a suitable model for the number of cherries in a *Rays* fruit cake.

(1)

The number of cherries in a *Rays* fruit cake follows a Poisson distribution with mean 1.5

A *Rays* fruit cake is to be selected at random.

Find the probability that it contains

(c) (i) exactly 2 cherries,

(ii) at least 1 cherry.

(4)

Rays fruit cakes are sold in packets of 5

(d) Show that the probability that there are more than 10 cherries, in total, in a randomly selected packet of *Rays* fruit cakes, is 0.1378 correct to 4 decimal places.

(3)

Twelve packets of *Rays* fruit cakes are selected at random.

(e) Find the probability that exactly 3 packets contain more than 10 cherries.

(3)

(Total for question = 14 marks)

Q22.

Paul takes the company bus to work. According to the bus timetable he should arrive at work at 08 31. Paul believes the bus is not reliable and often arrives late. Paul decides to test the arrival time of the bus and carries out a survey. He records the values of the random variable

X = number of minutes after 08 31 when the bus arrives.

His results are summarised below.

$$n = 15 \qquad \sum x = 60 \qquad \sum x^2 = 1946$$

(a) Calculate unbiased estimates of the mean, μ, and the variance of X.

(3)

Using the mean of Paul's sample and given $X \sim N(\mu, 10^2)$

(b) (i) calculate a 95% confidence interval for the mean arrival time at work for this company bus.

(ii) State an assumption you made about the values in the sample obtained by Paul.

(5)

(c) Comment on Paul's belief. Justify your answer.

(2)

(Total for question = 10 marks)

Examiner's Report

Q1.

In part (a) a large number of students ignored the outlier and gave an answer of 30 instead of 39. Others failed to score the mark because they left their answer as $48 - 9$ or they gave two answers: one with and one without the outlier. Part (b) on the other hand was answered very well with most students scoring the mark. Part (c) caused problems for some students who did not use a correct lower class boundary with 64.5 being a common error. Some of these students managed to correct this error in part (d) and most were successful here. The given answer tempted some to take the midpoint of the 60-65 interval and round up to get 63. We allow students to use $(n + 1)$ instead of n when calculating the cumulative frequency to use in the interpolation. Most used 9.5 correctly in part (c) but in part (d) few choosing this approach used the correct value of 15.25. In part (e)(i) many students were able to find the correct limits of 93 and 45 and make a suitable comment about the outliers. A sizeable minority though used the interquartile range from part (b) instead of the correct value of 12 and lost these marks. The box and whisker plots were usually correct with only a few making careless errors such as plotting the top whisker at 85 instead of 84. The final part proved to be a good discriminator with many revealing a lack of understanding about how to interpret the statistics they had calculated. Most achieved two of the 3 marks for choosing the 70° angle and giving a suitable reason based on the closeness of the median to this value or the smaller range or inter quartile range but few gave both reasons. Some surmised that a value in the range 60-65 would be more accurate based on this being the modal class, others referred to the absence of outliers or tried to use symmetry to justify their choice.

Q2.

Most students had a thorough understanding of box-plots and skewness and there were many fully correct answers. In question (a) the correct values were usually given but $a = 40$ and $b = 73$ were common incorrect answers where students failed to read the "leaves" from right to left .

In question (b) most students calculated the upper limit of 91 for outliers correctly and there were few errors in reading the scale and plotting 25, Q_1, Q_2 and Q_3 correctly. The upper whisker did cause some problems. An upper whisker stopping at 91 (the upper limit for outliers) or stopping at 75 (the next non-outlier in the data set) and then an outlier clearly shown by a cross or $*$ at 99 was all that was required. Common errors were to draw both whiskers or to plot the outlier at 91 or simply to extend the whisker to 99 and indicate the outlier there too.

In question (c) most comments about the skewness of Penville were correct but some did not realise that they needed to comment on Greenslax too. The reasons were usually based on a comparison of the differences in the quartiles and were often correct although a number still think that $Q_3 - Q_2 > Q_2 - Q_1$ means negative skewness. Some students elected to calculate the means to compare with the medians, which was accepted if the means were correct, but others tried an argument based on the mode failing to realise that the distributions were bimodal and therefore this approach was invalid.

Q3.

Part (a) was generally well answered, with most candidates getting the correct values. Only a few stated 7 and 1 (frequency density) and a few quoted other figures. A simple check that the total number of motorists should add to 100 would have told candidates if their calculations were incorrect. Some could not cope with the different interval widths and/or intervals starting and ending in ".5". This was more evident in part (b) where many struggled. The most common error was simply adding 21 and 45 not realising that 13.5 took them into the next group. A large number of candidates did realise that they needed to add

three different frequencies together, although some found it difficult to get the correct fraction of the third class. Some ignored the third class or used its whole frequency.

Q4.

Most could find the correct width in part (a) but the height calculation is still causing problems. The common error was to compare frequencies rather than frequency densities and an incorrect answer of 3 was fairly common. Part (b) was answered well by the majority of students with very few cases of students using (n + 1) and most choosing to work upwards from 45 rather than down from 50. The interpolation technique is understood but there were still a number of errors made in determining the appropriate class boundaries with 44.5 being used instead of 45 quite often.

As expected, part (c) was answered very well by the majority of students but there were still some who simply added the 5 midpoints and divided by 5 which is disappointing for students at this level. Many students didn't seem to be familiar with the convention that answers involving money should be given to
2 decimal places and some lost a mark for simply writing down £49.1 without any correct working being seen. There were many fully correct answers to part (d) as well, but some forgot to use the square root and a number lost accuracy by using 49.1 or the mean rather than the full expression.

In part (e), all that was required was a simple comparison of Q_2 with $\overline{x}$ but a number used a formula for skewness and that was, of course fine. The expected conclusion was that there was very little skewness since $Q_2 \approx \overline{x}$ but many concluded that since $\overline{x} > Q_2$ there was positive skew and this was allowed. The idea of using a normal distribution to model these data being an acceptable approach because there was only a little skewness was not appreciated by many students who simply trotted out a standard response that a normal was not a suitable model because the data was skew. Students should be encouraged to consider the magnitude of the skewness and realise that, with such a small amount of skew, to model the situation with a normal distribution may be a reasonable thing to do. Some students completely missed the need to reference part (e) in their answer to part (f) and simply compared their mean and standard deviation to Rika's values of £50 and £10. The final part (g) defeated many students despite being a fairly standard normal distribution question, presumably because it was not part of the question on the normal distribution. The three common errors were: standardising with their mean and standard deviation rather than the given values of 50 and 10, equating $\frac{c-50}{10}$ to a probability rather than a z-value or using 0.84 rather than 0.8416 from the percentage points tables.

Q5.

There were many incorrect answers of 38 given to part (a) where candidates, presumably, thought the question wanted the value that 75% of candidates exceeded rather than the value exceeded by 75% of the candidates. They should be encouraged to read the question carefully.

Part (b) was answered very well and where they showed some calculations to determine the outliers they usually scored well on part (c) too. A surprisingly large number though failed to show any working for the outliers and their diagram showed no evidence of them either. There was still some uncertainty about where to end the whiskers (either at the outlier limits or at the next non-outlier value are acceptable) and there was some evidence that candidates had been using previous mark schemes as "model solutions" and were drawing both sets of whiskers which is, of course, incorrect. There were many good answers to part (d) but some candidates failed to secure marks as they did not use the appropriate terms. The use of "mean" or "average" or "spread" are not sufficiently clear: we require use of the correct terms such as "median", "range" or "Inter Quartile Range".

Q6.

Again the majority of the candidates encountered few problems here and many scored 3 or 4 marks, although a number in part (a) found both means and then simply added them and divided by 2 rather than taking the weighted average. Most chose to find the sum of all 28 records and then simply divided this by 28. A few candidates misinterpreted the 84.6 as the mean for the 21 days despite what should have been an obvious discrepancy with Keith's data.

In part (b) many realised that the changes would have no effect on the mean but sometimes they failed to give any numerical values to support this claim. There were many excellent answers though showing clearly that 9.4 + 0.5 = 4.9 + 5.0 or some other suitable calculation.

Q7.

There seemed to be a slightly improved understanding of the calculation of standard deviation this summer and many correct responses to part (a) were seen. The usual errors still occurred though: forgetting the square root, or failing to divide the 985.88 by 8 or simply confusing S_{xx} with the variance.

The responses to part (b) confirmed that few students are aware of the simple relationship between S_{yy} and standard deviation as many attempted to find $\sum y^2$ by "unpacking" the value for δ_y with varying degrees of success. The calculation for S_{xy} in part (c) was carried out successfully by most students and many completed part (d) correctly too, usually by first finding Sxx rather than using their answer to
part (a).

Responses to part (e) were very mixed. Some simply quoted a statement about coding not affecting correlation and therefore expected no change. Others realised that the new point was breaking the trend (both values above their means conflicting with the negative correlation) but were not precise enough about the effect on the value of the correlation coefficient. The examiners were looking for a statement about the value that told us that the correlation was weaker, i.e. that the value was closer to zero:
r "increases" or r is "less negative" were not deemed to be satisfactory.

Q8.

In part (a) some candidates could not calculate the widths of the intervals and therefore lost all the marks. In part (b) the technique of linear interpolation is understood well but a number of candidates could not find the correct end-points. Candidates should look carefully at tables of grouped data and determine carefully the end points and widths of the intervals.

Parts (c) and (d) were answered well without quite so many false attempts at standard deviation as is often the case on S1. Part (e) was not answered so well as many candidates didn't appreciate the need to interpolate. Those who did usually arrived at the correct answer quite efficiently.

Q9.

There were many fully correct box plots in part (a) that gained full marks but also a number of almost correct plots with no supporting working that scored very few marks. Those who did show their calculations for determining the outliers usually got the 24.5 value correct but a large number had 3.5 rather than − 3.5 as their lower limit. Some candidates drew two upper whiskers one ending at 20 (the next highest non-outlier data point) and one ending at 24.5 (the outlier limit). A correct answer should only have one whisker and it appears that some candidates had been copying the practice on the mark scheme of illustrating both alternatives for the benefit of the examiners. It should be remembered that the published mark schemes are <u>not</u> model solutions and they should not be offered to candidates as such.

In part (b) most candidates stated that $Q_2 - Q_1 > Q_3 - Q_2$ or something equivalent though some then claimed that the skewness was positive but this was usually answered well.

Part (c) caught many candidates out with a large number agreeing with the company's claim because
Q_3 = £14 000 which is greater than £10 000. Those who did appreciate that the lower quartile was the significant figure to be considering often gave excellent answers but the success rate here was lower than expected.

Q10.

A small number of candidates still failed to calculate the mean correctly in part (a). For some this was due to errors with the midpoints but the more extreme errors involved dividing by 6 rather than 200 or using the class widths rather than the mid-points. The standard deviation formula still causes problems for some: forgetting the square root and failing to divide $\sum f t^2$ by 200 were common errors and some candidates used their rounded value of the mean and lost the final accuracy mark as their answer was not accurate to 3 significant figures. The calculation of the median in part (b) was answered well but applying the same principles to part (c) caused difficulties for some with many of those attempting the $(n + 1)$ approach using 50.5 instead of 50.25 and others using incorrect end points. In part (d) a few spotted that Q_3 was on the class boundary and gave the value of 25.5 but others encountered similar problems to those with Q_1 but most were able to find their interquartile range. Part (e) was not answered well with many mentioning "continuous data" or "extreme values" and only a few stating that their data was skewed. Most candidates scored some marks in part (f) but many failed to secure all the marks because they did not deal with all of the estimates; in particular the standard deviation was often omitted.

Q11.

The first three parts were generally completed with confidence and fluency. Only a few candidates found incorrect values for the median (as they did $\frac{45}{2}$ and then looked for the 22.5th value) or the quartiles. The standard deviation was particularly well answered, but the usual errors of not dividing sum of x^2 by 45 or forgetting the square root were the most common.

In part (d) most candidates knew a rule to apply but a few got muddled with what exactly it was, using
$Q_3 - Q_1$ or $Q_1 - Q_2$ or incorrectly for example. Having managed to apply the rule, some then got muddled in their reasoning, stating it showed positive skew or put the inequality signs the wrong way around which led to an incorrect conclusion. Those who simply used mean < median tended to fare better.

Part (e) was poorly answered, with some candidates not even attempting an answer. Perhaps they did not realise that this question considered a 'new' data set meaning it had

little to do with previous work, with many performing calculations on the sample of 45 rather than considering the population. There were many scripts containing long and elaborate wrong calculations. The scaled mean was done better than the scaled standard deviation. On finding the new mean many candidates correctly subtracted 5 but multiplied their answer by 0.1 and did not subtract. A large number of those who attempted decoding applied the same rule to both the mean and standard deviation or stated incorrectly that the standard deviation was not affected by coding. Only a few candidates had intrinsic understanding that a measure of spread was only affected by the 'multiplier'. Those candidates who knew what to do produced short, efficient solutions.

Q12.

The comment requirements in parts (e) and (f) made the latter stages of this question quite demanding but many were able to pick up marks in parts (b), (c) and (d). In part (a) most identified the width as
0.5 cm but, as usual, the height caused some difficulties. A number of students still seem to think that the height represents the frequency and some managed to make the statement that 4 cm^2 represents
8 babies but were still unable to use this to find the correct height.

The linear interpolation in part (b) was usually carried out very well but a few used the midpoints or made errors with class boundaries or used incorrect cumulative frequencies such as 10 or 27 instead of 9
and 26.

In part (c)(i), some students did not show how to use the midpoints to find $\sum fx$, simply using the value of 171.5; a "show that" question always requires a full explanation of the method used. More students now seem to know a suitable formula for standard deviation but some are still a little uncertain about whether or not a square root is required and some wasted time by calculating $\sum fx^2$ (not always successfully) rather than using the value given in the question.

Part (d) was a simple normal distribution probability calculation and the success rate amongst students with this standard application has improved considerably. A small minority are still uncertain about whether or not to subtract the value from the tables from 1 but there were many fully correct responses to this part

In part (e), a number made a correct observation about the mean and median and possible skewness and could relate this to the suitability of using the normal distribution. Far fewer, though, were able to compare their answer to part (d) with the equivalent proportion from the table and so missed out on the other two marks. In part (f), most stated that the mean would not change and often gave a correct reason for this but fewer knew that the standard deviation would decrease and fewer still were able to give a convincing reason to support this.

Q13.

Most students secured the mark in question (a) although a few gave the boundaries for the whole interval. Question (b) was seen to be challenging. Many students were not confident with the calculations for the height of bars on a histogram and in this case some could not deduce which bar was the second tallest. In question (c) some found the correct fraction of the upper or lower group but often they failed to do this for both groups. Most knew the correct technique to tackle question (d) but errors often occurred with the lower class boundaries or when rounding. Question (e) was answered well with most using their calculated quartiles to justify their answer although a few calculated the mean and compared this to the median.

Q14.

In (a) there is still some uncertainty about the concept of histograms. Most candidates could find the width of the required bar but a number still failed to consider frequency densities to find the height. Some attempts were on the right lines but used the interval [240, 320); others assumed height was linked to frequency alone and 4.5 cm was a common incorrect answer. There has been a clear improvement in the use of interpolation to calculate the median in recent sessions and many had $\frac{10}{22} \times 80$ (with only a small minority using 10.5 instead of 10). The common error was to use 239.5 rather than 240 as the class boundary but there were many fully correct answers seen.

The calculation of the mean in (c) was usually correct but the standard deviation still causes difficulties for some who failed to use a correct formula.(d) was straightforward although some failed to secure the second mark because they did not give a description (e.g. "positive") of their skewness.

In (e) the insertion of a normal calculation in this context seemed to fox some candidates but many were able to carry out this routine calculation correctly. However the interpretation in (f) was not answered so well. Some drew the connection between their answer to (e) and the proportion in the table and deduced that a normal distribution was reasonable (over this range). Slightly more focussed on the fact that in (d) they had stated that the distribution was skewed and therefore a normal distribution was probably not a suitable model (overall) but few were able to marry these statements up and secure both marks.

Q15.

This question was not answered very well. Many candidates either made a poor attempt at part (a) and then abandoned the question or just left it blank and moved on. Those who correctly formed a frequency table often scored well, whilst others who failed to complete part (a) struggled to make any headway with the remainder of the question.

In part (a) there needed to be some attempt to count squares and 22.5, 562.5 or 112.5 (small squares greater than or equal to 35 mph) were frequently seen. However many candidates did not appreciate the key idea that area is proportional to frequency and there was no attempt to combine this figure with the total frequency of 450. Those who did combine their figures and were able to come up with a correct relationship between area and number of cars (e.g. 1 large square represents 20 cars) were usually able to complete this part successfully although a few found the number of cars speeding above 30 mph instead of 35 mph as required. A few candidates stumbled upon 90 by dividing the 450 cars by the 5 bars of the histogram but this, of course, received no credit.

In part (b) most attempts tried to use mid-points but many struggled to find suitable frequencies and a few were unsure of the class widths (using 6 and 11). Some used the number of squares as frequencies but they rarely had a compatible denominator for their expression for the mean.

Most attempts at part (c) realised that interpolation was required but many promising solutions used 19.5 or 20.5 as class boundaries rather than 20.

Although they may not have had the correct values for the mean and median many had some values which they could use to answer part (d). A simple comparison of their values (e.g. mean greater than median) earned them the first mark and then an appropriate comment about the skewness (such as positive skew) the second. Some attempted to calculate the quartiles and invariably these were incorrect. Candidates should consider the amount of work involved in finding these values and compare it with the tariff for the

question: 2 marks for a comment and a reason should not involve half a page of calculations. Other candidates tried to justify their comment from the shape of the histogram ignoring the calculations in (b) and (c).

In part (e) we required a choice of mean or median that was compatible with their conclusion in (d). Some candidates who had correctly concluded that the distribution was skewed in (d) still chose the mean, because it uses all the data, but there were many correct answers seen to this part.

Q16.

Part (a) was, unsurprisingly, answered very well and most candidates were able to use the mid-points to find the mean in part (c). The calculation of standard deviation is still causing problems though: an incorrect formula, premature rounding and forgetting the square root were the typical errors. The technique of interpolation required in part (b) is now better understood but incorrect class widths still caused problems here. A common response to part (d) was to simply state that the data was continuous rather than using the calculations they have just performed in parts (b) and (c).

Part (e) was usually answered well with most using $Q_3 - Q_2 > Q_2 - Q_1$ but a small minority gave the wrong conclusion despite giving the correct comparison.

Q17.

The use of interpolation, which was expected in part (a), is improving and many candidates made a reasonable attempt. Some failed to use correct class limits and had a width of 9 rather than 10 but the correct answer was often seen. Despite the values for Σt and Σt^2 being given in the question a number of candidates chose to estimate these values from the table and subsequently obtained an incorrect estimate for the mean but the mark scheme did permit them to obtain the remaining marks. The question did not ask for estimates of the mean and standard deviation and so using the mid-points was inappropriate in this question. The usual crop of errors arose when calculating the standard deviation with many failing to divide the 69378 by 32. Candidates can usually calculate S_{xx} correctly using the given formula and they may find it helpful to simply remember that standard deviation is simply $\sqrt{\frac{S_{xx}}{n}}$ but few seem to use this approach.

In part (c) most compared the mean and median and correctly concluded that the skewness was positive. Some used a formula which, apart from the extra work, was fine but those who used a (median - mean) formula rarely gave the correct conclusion. Some candidates went to great lengths to calculate the quartiles and conducted a quartile test for skewness which, if correct, led to a conclusion of negative skewness. This was allowed but was clearly not the intended approach for 2 marks.

Q18.

Finding the midpoints of the given groups was predominantly carried out correctly with very few errors seen. In contrast, attempts at finding the width and height of the 26-30 group were extremely varied, with most candidates finding this particularly challenging, especially in finding the height. In the majority of cases, candidates obtained the wrong width and height, mostly with no clear strategy, although these did multiply together to make 20.8 in some cases. Calculation of the mean was carried out successfully on the whole, although there were some apparent misconceptions, with quite a few candidates merely summing the midpoints (without multiplying by the frequency) and dividing this by 56.

The standard deviation proved to be more problematic, with frequent mistakes in both the formula and in their calculations. Some candidates used the sum of the f^2x's and others the sum of the $(fx)^2$ or the sum of the fx's all squared in their formula. Very few candidates calculated s. Most candidates were able to use the correct interpolation technique to obtain the median, although many lost the accuracy mark through their use of 21 as the lower class boundary (which was relatively common) and /or 4 as the class width. Quite a few candidates worked with 28.5. A few candidates tried to apply the correct formula to the wrong class interval, however. Some candidates appeared to have a limited understanding of the class boundaries and failed to recognise the continuous nature of the data.

The majority of candidates were able to carry out a suitable test to determine the skewness of the data correctly. This mostly involved comparing $Q_3 - Q_2$ to $Q_2 - Q_1$ (with or without explicit substitutions), although the wrong conclusion was often drawn, either following on from a previous error in evaluating the median or from a lack of understanding of what their result was showing. A few students evaluated 3(mean-median)/standard deviation. Quite often the result of their test was described in words not figures, for example Q_2 is closer to Q_3 than Q_1. Some candidates merely attempted to describe the skewness without carrying out any test.

Q19.

Many students are still unsure about calculating the heights of bars in histograms but the work on linear interpolation and even standard deviation seems to be improving.

Most were able to state the correct width of the bar but few used frequency densities correctly to find the height, some finding the frequency density of $\frac{1}{3}$ but then calculating $\frac{1}{3} \times 2.5$ rather than $2.5 \div \frac{1}{3}$. Some identified that 2 1.5 cm represented 10 customers but were then unable to use this correctly to find the height. Question (b) was answered well but some students had an incorrect class width because they did not realize that the lower class boundary was 70 not 69.5. question (c) the mean was usually correct, although a few weaker students divided 6460 by 6 (the number of groups), and there were fewer errors made in calculating the standard deviation: some forgot the square root and others forgot to divide $\sum fx^2$ by 85. A few students ignored these given values and recalculated $\sum fx$ and $\sum fx^2$ (often incorrectly!) and wasted valuable time. In question (d) most were able to use the given formula with their values and most gave a sensible comment based on their evaluation.

Q20.

This question proved to be quite challenging for a high proportion of candidates. A significant number either made no attempt at the question or offered very little in the way of creditable solutions, with many unable to progress beyond part (a). Time issues may have been a contributing factor in some cases.

The majority of candidates however, were able to earn some credit at least in part (a), for their standardisation, although whilst this was often completely correct, a fairly common mistake was to give $1-0.8944 = 0.1056$ as their final answer.

Many students did not recognise that they needed to actually use the normal distribution in part (b) and part (c), giving rise to extremely poor attempts by numerous candidates. Of these, many merely gave 45 and 15 as their quartiles, whilst others calculated ¾ of some value as their upper quartile (for example ¾ × 60) and ¼ of the same value as their lower quartile. Alternatively, of those who understood that they were required to use the normal distribution, most attempts were successful, though there were some instances of their setting their standardisation equal to a probability, usually 0.75 or $P(Z < 0.75)$, and not a z-value. Unfortunately 0.68 was used fairly frequently as the z value. The majority of candidates were however able to follow through their value of the upper quartile to find their lower quartile using symmetry, though some performed a second calculation involving standardisation. Some candidates miscalculated their lower quartile as 1/3 of their upper quartile.

Despite previous errors most candidates tended to be successful in substituting their values correctly into at least one of the given formulae. However, a few seemed unaware of the order of the operations.

The final part of the question also proved difficult for many candidates with some running into trouble as a consequence of previous errors in part (b), part (c) and part (d) and others providing no attempt at all. Indeed, for numerous candidates, incorrect values for h and k led to probabilities of 0 being calculated from results such as $P(Z > 7)$ and thus many creditable attempts lost marks through earlier inaccuracies.

Q21.

Part(a) of this question was poorly answered. A significant proportion of candidates were unable to find the mean and variance of a frequency distribution in part (a). Almost all candidates obtained the sums of fx and fx^2. However, some candidates did not perform any division. Other candidates divided by seven.

In part (b). there were some candidates who were aware why the evidence in this case (mean ≈ variance) suggests that a Poisson distribution may be appropriate. However, many candidates chose instead to mention the conditions under which one might consider a Poisson distribution in the first place, without any knowledge of any actual data.

Part (c), however, was a good source of marks for a very large number of candidates. Even those who had lost the first four marks of the question were able to respond to the challenge of S2 material by recalling the knowledge and performing the skills required by the question.

It was pleasing that almost all candidates provided full detail of the method and working for part (d): this approach is always advisable, but essential when the final answer is given in the question.

Part (e) was also a good source of marks for the vast majority of candidates. A minority of candidates used a Binomial distribution with $n = 10$.

Q22.

Most students were able to gain the marks for part (a). Very few found the sample variance without multiplying by $(n+1)/n$.

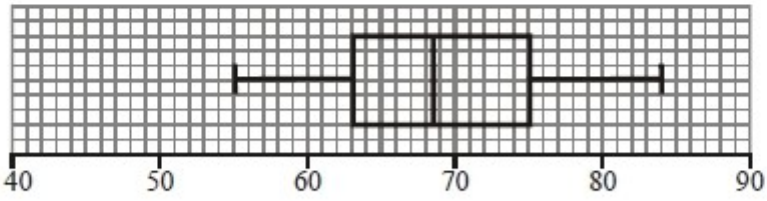
In part (b) it was most common for students to gain the first three marks here for a correct attempt at the confidence interval, with very few not using 1.96 for the z-value. A small number lost an accuracy mark for using their s instead of the stated standard deviation of 10. It was very common however for the CI not be stated in terms of the arrival time as required in the question, with most leaving it in terms of minutes late, hence missing the final mark in part (i). Part (ii) was not well answered with few giving assumptions about the bus arrival times. Often they just referred to 'values' being independent with no context. A number of students did not understand what they were being asked in this context and stated various textbook answers referring to the Central Limit Theorem.

Part (c) was the least successful part of the question with only a minority explicitly stating that the stated arrival time was in the interval. Some trivially stated that the mean was in the interval. Some students focussed solely on the bottom end of the interval being negative or below the stated time, without stating the time was contained in the interval. Many students felt that the most significant feature was that the time was near the lower limit so that even if they did say it was in the interval they typically decided that Paul was correct and that the bus was more often late i.e. they did not demonstrate an understanding of the meaning of a confidence interval.

Mark Scheme

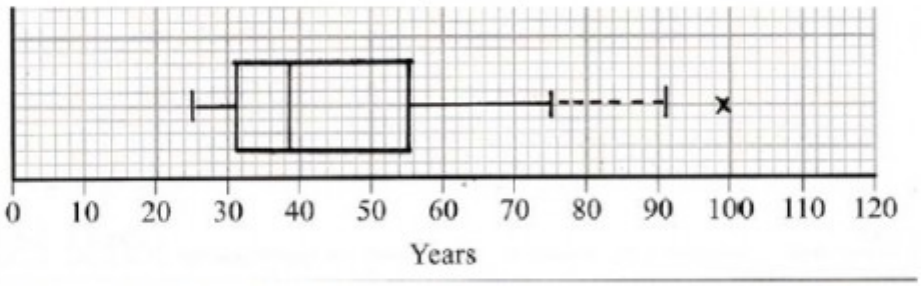
Section 1	Numerical measures, graphs and diagrams
qu.1	June 2015 S1 - Q1
qu.2	June 2014 S1 - Q1
qu.3	January 2012 S1 - Q1
qu.4	June 2017 S1 - Q2abcdg
qu.5	May 2013 S1 - Q2
qu.6	January 2011 S1 - Q2
qu.7	June 2016 S1 - Q3a
qu.8	May 2013 S1R - Q3abc
qu.9	January 2011 S1 - Q3
qu.10	May 2013 S1 - Q4 all (f*)
qu.11	January 2012 S1 - Q4abce
qu.12	June 2016 S1 - Q5abcf*
qu.13	June 2014 S1R - Q5
qu.14	January 2013 S1 - Q5abc
qu.15	May 2012 S1 - Q5 all (e*)
qu.16	May 2011 S1 - Q5ab
qu.17	January 2011 S1 - Q5ab
qu.18	May 2010 S1 - Q5abcd
qu.19	June 2014 S1 - Q6abc
qu.20	May 2010 S1 - Q7
qu.21	June 2016 S2 - Q1a
qu.22	May 2017 S3 - Q5a

Q1.

Question	Scheme	Marks
(a)	[Range = 48 − 9] = **39**	B1 (1)
(b)	[IQR = 25 − 12]= **13**	B1 (1)
(c)	Median = $65+\frac{[9]}{13}\times 5=\frac{890}{13}=$ awrt **68.5°** $\left[\text{Condone: } 65+\frac{[9.5]}{13}\times 5=68.7\right]$	M1 A1 (2)
(d)	Lower Quartile = $60+\frac{9}{15}\times 5=$ **63** (*)	M1 A1cso (2)
(e)(i)	$63-1.5\times(75-63)=45$ $75+1.5\times(75-63)=93$ No data above 93 and no data below 45 or 55>45 etc or there are no outliers.	M1A1 A1
(ii)		M1 A1ft (5)
(f)	Median for the 70°angle is closer (to 70°)[than the 20° median is to 20°] The range/IQR for the 70° angle box plot is smaller/shorter Therefore, students were more accurate at drawing the 70° angle.	B1 B1 dB1 (3)
		(14 marks)

	Notes
(c)	M1 for an attempt (should have 65 or 70, 13 and 5)NB working down: $70-\frac{[4]}{13}\times 5$ Allow any correct method leading to $\frac{890}{13}$, the "5" may be implied by 65 and 70 seen A1 awrt 68.5 (condone 68.7 if (n+1) is used). Ans only of 68.5 is 2/2 but 68.7 needs M1
(d)	M1 for correct expression for the lower quartile (condone 9.25 if (n+1) used) Watch out for working down e.g. $65-\frac{6}{15}\times 5$ (M1) but e.g. $\frac{60+65}{2}=62.5=63$ is M0 A1 for correct solution with no incorrect working seen (condone (n+1) giving 63.08..)
(e)(i)	M1 for either correct calculation (may be implied by one correct limit) A1 for either 45 or 93 A1 for 45 and 93 and conclusion
(ii)	M1 for a box with 1 whisker drawn on each side (must see the line drawn) A1ft their median $63 < Q_2 < 75$ but quartiles (63 and 75), 55 and 84 must be correct.
Accuracy	Use 0.5 sq. accuracy so condone median on 68 or 69 if 68.5 seen
(f)	1st B1 for correct comparison of their **medians** (63 < (c) < 75) to true value 2nd B1 for correct comparison of their **range** or **IQR** ("spread" is B0) Allow saying IQRs of 12 and 13 are similar. Ignore mention of "skewness" or "outliers" 3rd dB1 dependent upon at least one previous B1 being scored for choosing 70°

Q2.

Question Number	Scheme	Marks
(a)	$a = 44$ $b = 76$ These answers may be in or near the table	B1 B1 **(2)**
(b)	$55+1.5(55-31)=91$ [and $31-1.5(55-31)=-5$] **Penville**  	M1 B1 B1 A1 **(4)**
(c)	Greenslax : $[Q_2-Q_1=20,\ Q_3-Q_2=12$ or $(Q_2-Q_1)>(Q_3-Q_2)] \Rightarrow$ –ve(skew) Penville: $[Q_2-Q_1=8,\ Q_3-Q_2=16$ or $(Q_3-Q_2)>(Q_2-Q_1)] \Rightarrow$ +ve (skew) Don't insist on seeing "skew" so just –ve and +ve will do. Treat "correlation" as ISW Justification that is consistent	B1 B1 ddB1 **(3)** **Total 9**

	Notes
(b)	**A fully correct box plot scores 4/4. If <u>not</u> fully correct apply scheme and need evidence for M1** **If two box plots are seen ignore the one for Greenslax. If not on graph paper M1 max for (b)** M1 for sight of $55+1.5(55-31)$ <u>or</u> 91 seen (possibly implied by RH whisker of box plot) May be implied by a fully correct box plot 1st B1 box with whiskers (condone missing median) 2nd B1 25, 31, 39, 55, RH whisker to end at 75 or 91. Two RH whiskers is B0 Accuracy must be to within 0.5 of a square so e.g. lower quartile at 30 or 32 is OK
(c)	A1 only one outlier plotted at 99. Allow cross to be vertically displaced If the RH whisker goes to 99 (2nd B0) <u>and</u> A0 even if outlier is identified since we require a horizontal "gap" between RH whisker and outlier. 1st B1 Greenslax – ve (skew) 2nd B1 Penville + ve (skew). We must be able to tell which is which but labels may be implied by their <u>values</u> but not simply from $Q_3-Q_2>Q_2-Q_1$ If there is just <u>one</u>, unlabelled comment assume Penville. 3rd ddB1 dependent on 1st and 2nd B marks being scored. Justification for <u>both</u> based on: quartiles, median relative to quartiles, or "tail" If only values for Q_3-Q_2 etc are <u>given</u> they should be correct ft for Greenslax and correct for Penville If values for Greenslax imply +ve skew then 1st B0 and 3rd B0

Q3.

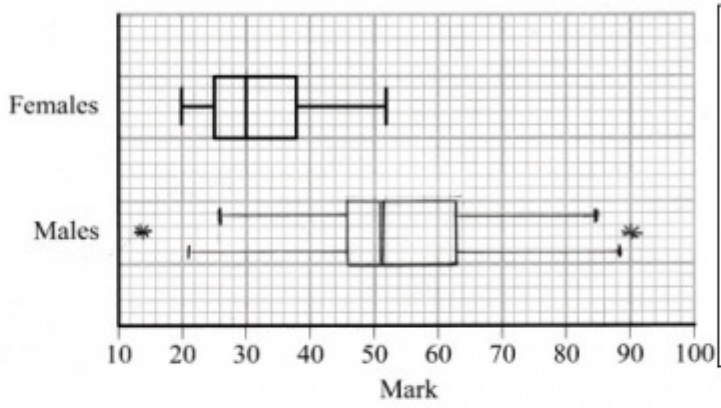
Question Number	Scheme	Marks
(a)	14, 5	**M1 A1** **(2)**
(b)	21 + 45 + 3 = 69	**M1 A1** **(2)** **Total 4**
NOTES		
(a)	M1 for 2x7 or 14 or 5x1 or 5	
	A1 for both 14 and 5	
(b)	M1 for 21+45+(0<frequency <9)	
	A1 for 69 only.	
	69 no working, award M1A1 Incorrect answer with no working M0A0	

Q4.

Question Number	Scheme	Marks
(a)	Width (w) = **4** cm	B1
	Areas: 16 cm^2 represents 32 offices (o.e.) or their $h = \frac{6}{\text{their } w}$ (3sf) or $\frac{8}{3.2}\times 0.6$	M1
	So height (h) = **1.5** cm	A1 (3)
(b)	e.g. $(45)+\frac{20}{25}\times 5$ or $(50)-\frac{5}{25}\times 5$ (o.e.); = (£) **49**	M1; A1 (2)
(c)	$\frac{\sum fy}{90}=\frac{4420}{90}$, = (£) **49.11** (or better) (Allow $\frac{442}{9}$ or $49\frac{1}{9}$)	M1, A1 (2)
(d)	$\sqrt{\frac{226687.5}{90}-\bar{x}^2}=\sqrt{106.8487...}$, = 10.3367 = awrt (£) **10.3**	M1, A1 (2)
(e)	Mean ≈ median so distribution is symmetric (no skew or very little skew) [Allow mean > median or $k(\bar{x}-Q_2)$ (k>0) so +ve skew if compatible with their figures] [If using quartiles we must see Q_1 = 44.0 and Q_3 = 55.5 used]	B1ft (1)
(f)	Symmetric (or little skew) so normal (or Rika's suggestion) may be suitable	B1ft (1)
(g)	$\frac{c-50}{10}=0.8416$ [N.B. use of (1 – 0.8416) is B0]	M1, B1
	c = 58.416 = (£) 58.42 awrt **58.4**	A1 (3) **[14]**

	Notes
(a)	M1 for a correct calculation of areas 1 cm^2 = 2 offices (o.e.) A1 for h = 1.5 cm (Correct answer only 2/2)
(b)	M1 for a correct expression without end point. Allow "n + 1" so e.g. $(45)+\frac{20.5}{25}\times 5$ A1 for 49 or, if (n + 1) used, allow 49.1 (Correct answer of 49 only 2/2)
(c)	M1 for an attempt at $\frac{\sum fy}{90}$ with at least 3 correct products of $\sum fy$ or $4000\leq\sum fy\leq 5000$ A1 for 49.11 (Allow 49.1 from correct working) (Correct answer only 2/2, 49.1 only M1A0)
(d)	M1 for a correct expression including $\sqrt{\ }$, ft their mean. Allow use of s A1 for awrt 10.3 Allow s = awrt 10.4 if clearly used. [NB use of 49.1 gives 10.389 ⇒ A0 (Correct answer of 10.3 with no working is 2/2)
(e)	B1ft for reason and "symmetric" (or other correct) statement [Allow positive skew] Allow ft of their (b) and their (c). For "symmetric" need $\lvert\bar{x}-Q_2\rvert<1$ "correlation" is B0
(f)	B1ft Suggest normal is or isn't suitable **with** suitable reason based on (e) or mean and med
(g)	M1 for stand'ing using "c", 50 and 10 and setting equal to $\pm z$ value where $0.84\leq z\leq 0.85$ B1 for using $z=\pm 0.8416$ or better (calc gives 0.8416212…) in standard' attempt e.g. $\sqrt{10}$ for 10 A1 for awrt 58.4 (accept 3sf here) (Ans only of awrt 58.4 is M1B0A1 but 58.416 or better is 3/3)

Q5.

Question	Scheme	Marks
(a)	25 (allow any x where $24 < x < 26$)	B1 **(1)**
(b)	**Q_2** (or median or m) = **51** **IQR** = 63 – 46 ,= **17** (or $Q_3 - Q_1 = 17$)	B1 M1, A1 **(3)**
(c)	Outliers given by $46 - 1.5 \times 17 = 20.5$ or $63 + 1.5 \times 17 = 88.5$ Outliers limits are **20.5** and **88.5**  Allow lower whisker to 20.5 and upper whisker to 88.5 Do **not** allow a mix of whiskers e.g 20.5 and 85 Do **not** allow both sets of whiskers	M1 A1 M1 A1ft B1 **(5)**
(d)	**Medians:** Median for females lower than males **IQR:** IQR for females smaller than males. Allow "lower/higher" but not "wider" **Range:** Range of females is less than males **Skewness:** Male and female marks are both positively skew Ignore other statements about average, spread, mean, st. Dev, variation, outliers etc	B1ft B1ft **(2)** **(11 marks)**

	Notes
	Mark (b) and (c) together BUT must see clear statement that median (or m or Q_2) = 51 and IQR = 17
(b)	M1 for 2 quartiles (at least one correct) and attempt to find the difference. Must see their 63 – their 46 A1 for 17 only. [Answer only of IQR= 17 scores M1A1]
(c)	**A fully correct box-plot (either version) with no supporting work scores 5/5. Otherwise:** 1st M1 for correct attempt to calc' at least one limit for outliers, ft their quartiles or IQR or award for sight of 20.5 or 88.5 1st A1 for identifying both limits of 20.5 and 88.5 2nd M1 for a box with an upper and a lower whisker(s) with at least 2 correct values (or correct ft) (condone no median marked) (condone 2 upper or 2 lower whiskers) 2nd A1ft for their 20.5 or 26 ,46, 51, 63 and 85 or their 88.5 in appropriate places and readable off their scale. Follow through their 20.5 and their 88.5 only, other values need to be correct If there are 2 upper or 2 lower whiskers A0 B1 for only 2 outliers appropriately marked at 14 and 90 Do not award if whiskers go beyond these values. **Apply ± 0.5 square accuracy for diagram** A box plot not on the graph paper can only score the 1st M1A1
(d)	**In (d) ft from their diagrams (if no diagram then use their values)** 1st B1ft for one correct comment comparing median, IQR , range or skewness 2nd B1ft for a second correct comment comparing median, IQR, range or skewness Do not allow contradictory statements

Q6.

Question Number	Scheme	Marks
(a)	$2.8 + 5.6 + 2.3 + 9.4 + 0.5 + 1.8 + 84.6 = 107$ mean = 107 / 28 (= 3.821…) **(awrt 3.8)**	M1 A1 (2)
(b)	It will have no effect since one is 4.5 under what it should be and the other is 4.5 above what it should be.	B1 dB1 (2) **[4]**
	Notes	
(a)	M1 for a clear attempt to add the two sums. Accept a full expression or $2.8 + 5.6 + \ldots + 84.6 = x$ where $100<x<110$ i.e. seeing at least two correct terms of Keith's and the 84.6 with a slip. A1 for awrt 3.8 (Condone 1 dp/2sf here since data is given to 1 dp or 2 sf) Accept $\frac{107}{28}$ or $3\frac{23}{28}$ or any exact equivalent **Correct answer implies M1A1**	
(b)	1st B1 for clearly stating that it will have no effect. ("roughly the same" is B0 B0) 2nd dB1 for a supporting reason that mentions the fact that the increase and decrease are the same and gives some numerical value(s) to support this. e.g. Sum of Keith's observations is still 22.4 (or mean is still 3.2) or Sum is still 107 or $9.4 - 4.9 = 5 - 0.5$ (o.e.) This second B1 is dependent on their saying there is no effect so B0B1 is not possible.	

Q7.

Question Number	Scheme	Marks
(a)	$[\sigma_x^2 =] \frac{985.88}{8} - \left(\frac{86.8}{8}\right)^2 = \frac{985.88}{8} - 10.85^2$	M1
	$\sigma_x = \sqrt{\frac{985.88}{8} - \left(\frac{86.8}{8}\right)^2} = \sqrt{123.235 - 117.7225} = \sqrt{5.5125}$ or $\sqrt{\frac{44.1}{8}}$	A1
	$= 2.3478\ldots =$ awrt **2.35**	A1
		(3)
(b)	$S_{yy} = 8 \times \sigma_y^2 = 716$ (3 sf) but may see	
	$1136.584 - \frac{58^2}{8}$ or $27628(.084168) - \frac{464^2}{8}$ or 716.08... (= 716 to 3 sf) (*)	B1cso
		(1)
(c)	$S_{xy} = 4900.5 - 58 \times 86.8$ or $4900.5 - \frac{86.8 \times 464}{8}$	M1
	$= -133.9$ (Allow -134)	A1
		(2)
(d)	$r = \frac{-133.9/8}{\sigma_x \times \sigma_y}$ or $\frac{-133.9}{\sqrt{44.1 \times 716}}$	M1
	= awrt -0.753 or -0.754	A1
		(2)
(e)	$r < 0$ means high sunshine and low rain; this is high sunshine high rain	B1
	[this is not in keeping with the trend so] r is closer to 0 or $\lvert r \rvert$ decreases	B1
		(2)
		[10 marks]

	Notes	
(a)	M1 for a correct expr' for st. dev or variance (ignore label)[may be implied by 2.35 or 5.5125]	
	1st A1 for a correct expression for st. dev (must have square root) can ignore label	
	2nd A1 for awrt 2.35 (allow $s = 2.5099\ldots$ or awrt 2.51). If they have $\sigma^2 = 2.35$ score A0 but condone no label	
(b)	B1cso for a correct expression or sight of at least 716.08… (NB limits: 716.00~716.16)	
	Do not allow verification. Beware circular arguments: $716 \rightarrow \Sigma y^2 \rightarrow$ expr' $\rightarrow 716$	
(c)	M1 for a correct expression for S_{xy} (NB $\Sigma y = 464$)	
	A1 for -133.9 or awrt -134 [No fractions] (Answer only 2/2)	
(d)	M1 for a correct expression for r (ft their values for S_{xy} and σ_x or S_{xx})[Allow ft of S_{yy}]	
	A1 for awrt -0.753 or -0.754 (Answer only 2/2)	
(e)	**If they do not have an answer to (d) or their value of r is > 0 or $\lvert r \rvert > 1$ score B0B0 here**	
	1st B1 for a suitable reason contradicting $r < 0$ e.g. new value is not in keeping with trend or both $14 > \bar{x}$ and $70 > \bar{y}$ or saying both above average. Allow for $-0.48 <$ new $r < -0.47$	
	2nd B1 for a correct statement about r getting closer to zero e.g. $\lvert r \rvert$ decreases	
	A comment that r decreases or r is smaller or r is "less negative" is B0	
	"r increases" is B0 unless they also say that it gets closer to 0	

Q8.

Question	Scheme	Marks
(a)	Width = 2×1.5 = **3 (cm)**	B1
	Area = $8\times1.5=12$ cm^2 Frequency = 24 so 1 cm^2 = 2 plants (o.e.)	M1
	Frequency of 12 corresponds to area of 6 so height = **2 (cm)**	A1 (3)
(b)	$[Q_2=]\ (5+)\ \frac{19}{24}\times5$ **or** (use of $(n+1)$) $(5+)\frac{19.5}{24}\times5$	M1
	$=8.9583\ldots$ **awrt 8.96** **or** $9.0625\ldots$ awrt 9.06	A1 (2)
(c)	$[\bar{x}=]\ \frac{755}{70}$ or **awrt 10.8**	B1
	$[\sigma_x=]\ \sqrt{\frac{12037.5}{70}-\bar{x}^2}\ =\sqrt{55.6326\ldots}$	M1A1ft
	= **awrt 7.46** (Accept s = awrt 7.51)	A1 (4)
(d)	$\bar{x}>Q_2$	B1ft
	So positive skew	dB1 (2)
(e)	$\bar{x}+\sigma\approx18.3$ so number of plants is e.g. $\frac{(25-"18.3")}{10}\times12\ (+4)$ (o.e.)	M1
	= 12.04 so **12** plants	A1 (2)
		[13]

	Notes
(a)	M1 for forming a relationship between area and no. of plants or their width × their height = 6 A1 for height of 2 (cm). Make sure the 2 refers to height and not plants!
(b)	M1 for a suitable fraction ×5 (ignore end points) A1 for awrt 8.96 (or $\frac{215}{24}$ or $8\frac{23}{24}$) or 9.06 (or $\frac{145}{16}$ or $9\frac{1}{16}$) if using $(n+1)$
(c)	B1 for a correct mean. Accept exact fraction or awrt 10.8 M1 for a correct expression for σ or σ^2. Condone mixed up labelling- ft their mean A1ft for a correct expression – ft their mean but must have square root A1 for awrt 7.46 (use of s = awrt 7.51). Condone correct working and answer called variance.
(d)	1st B1ft for a correct comparison of their $\bar{x}$ and their Q_2
ALT	Allow use of a formula for skewness that involves $(\bar{x}-Q_2)$ or use of quartiles but must have correct values NB $Q_1=5.31$, $Q_3=14.46$ (awrt 14.5), $Q_3-Q_2\approx5.5$, $Q_2-Q_1\approx3.7/6$ 2nd dB1 Dependent on a suitable reason for concluding "positive skew" . "correlation" is B0
(e)	M1 for a suitable expression involving some interpolation (condone missing 4 so accept awrt 8) Condone use of end points of 25.5 and 14.5 in their interpolation expressions. A1 for 12 (condone awrt 12). Answer only 2/2

Q9.

Question Number	Scheme	Marks
(a)	Outliers $14 + 1.5\times(14-7) = 24.5$ $7 - 1.5\times(14-7) = -3.5$ Outlier 25 either upper limit acceptable on diagram 	M1 A1 M1 A1ft B1 (5)
(b)	Since $Q_3 - Q_2 < Q_2 - Q_1$. Allow written explanation negatively skew	B1 dB1 (2)
(c)	not true since the lower quartile is 7000 and therefore 75% above 7000 not 10000 or 10 is inside the box or any other sensible comment	B1 dB1 (2) [9]

	Notes
(a)	**A fully correct box-plot (either version) with no supporting work scores 5/5. Otherwise read on** 1st M1 for at least one correct calculation seen 1st A1 for 24.5 and −3.5 (or just negative noted) seen. May be read off the graph. If both values are seen but no calculation is given then M1A1,one value M1A0. 2nd M1 for a box with an upper and a lower whisker(s) with at least 2 correct values (condone no median marked) 2nd A1ft for 3,7, 12, 14 and 20 or 24.5 in appropriate places and readable off their scale If both upper whiskers are seen A0 Apply ft for their whiskers being compatible with their outlier limits e.g. if their lower limit is + 3.5 then a lower whisker ending at 4 or 3.5 is OK B1 for only one outlier appropriately marked at 25 **Apply ± 0.5 square accuracy for diagram**
(b)	1st B1 for $Q_3 - Q_2 < Q_2 - Q_1$ statement or an equivalent statement in words Use of $Q_3 - Q_2 < Q_2 - Q_1$ does not require differences to be seen. 2nd dB1 for "negative skew" dependent on suitable reason given above. "correlation" is B0 "positive skew" with a supporting argument based on whiskers can score B1B1 e.g. "right hand whisker is longer than LH one so positive skew" $Q_3 - Q_2 < Q_2 - Q_1$ followed by "positive skew" is B1B0
(c)	1st B1 for rejecting the company's claim 2nd dB1 for an appropriate supporting reason. Dependent on rejecting company's claim.

Q10.

Question	Scheme	Marks
(a)	$\sum ft = 4837.5$ (allow 4838 or 4840)	B1
	$\text{Mean} = \frac{"4837.5"}{200} = 24.1875$ awrt **24.2** or $\frac{387}{16}$	M1 A1
	$\sigma = \sqrt{\frac{134281.25}{200} - \left(\frac{4837.5}{200}\right)^2}$	M1
	$= 9.293$ ..... (accept s =9.32) awrt **9.29**	A1
		(5)
(b)	$Q_2 = [20.5] + \frac{(100/100.5 - 62)}{88} \times 5 = 22.659...$ awrt **22.7**	M1 A1
		(2)
(c)	$Q_1 = 10.5 + \frac{(50/50.25)}{62} \times 10 [= 18.56]$ (*) (n + 1 gives 18.604…)	B1 cso
		(1)
(d)	Q_3= 25.5 (Use of n + 1 gives 25.734…)	B1
	IQR = 6.9 (Use of n + 1 gives 7.1)	B1 ft
		(2)
(e)	The data is skewed (condone "negative skew")	B1
		(1)
(f)	Mean decreases and st. dev. remains the same. [Must mention mean and st. dev.] (from(a))	B1
	The median and quartiles would decrease. [Must refer to median and at least Q_1.] ((b)(c))	B1
	The IQR would remain unchanged (from (d))	B1 **(3)**
		(14 marks)

	Notes
	Correct answers only score full marks in each part except (c)
(a)	B1 for 4837.5 or 4838 or 4840 seen.
	If no $\sum ft$ seen (or attempt at $\sum ft$ seen), B1 can be implied by a correct mean of awrt 24.2
	1st M1 for attempt at their $\frac{\sum ft}{\sum f}$ allow 1sf so $\sum f$ = awrt 200 and $\sum ft$ = awrt 5000.
	Or award M1 for a clear attempt at mean where at least 4 correct products of $\sum ft$ are seen
	2nd M1 for correct expression including square root seen. Follow through their mean. Allow a transcription error in 134281.25 but not an incorrect re-calculation.
(b)	M1 for a correct fraction ×5. Ignore end point but must be +. Allow use of (n + 1) giving 100.5…
(c)	B1cso for a fully correct expression including end point. NB Answer is given. Allow use of (n + 1) giving 50.25…but use of 50.5 scores B0
(d)	1st B1 for 25.5 (or awrt 25.7 using n + 1)
	2nd B1ft for their Q_3 − their Q_1 (or 18.6) (provided > 0) Accept awrt 2sf. Correct ans. only scores 2/2
(e)	B1 Must mention that the data is skewed or not symmetrical. Do not award for "outliers"
(f)	1st B1 for one correct comment from the above. May refer to parts (a), (b), (c) or (d)
	2nd B1 for two correct comments from the above
	3rd B1 for all 3 correct comments from the above

Q11.

Question Number	Scheme	Marks
(a)	60	B1 (1)
(b)	$Q_1 = 46$ $Q_2 = 56$ $Q_3 = 64$	B1 B1 B1 (3)
(c)	mean = 55.48…. or $\frac{2497}{45}$ awrt 55.5	B1
	$sd = \sqrt{\frac{143369}{45} - \left(\frac{2497}{45}\right)^2}$	M1
	= 10.342… (s = 10.459..) anything which rounds to 10.3 (or s = 10.5)	A1 (3)
(d)	Mean < median < mode or $Q_2 - Q_1 > Q_3 - Q_2$ with or without their numbers or median closer to upper quartile (than lower quartile) or (mean-median)/sd <0;	B1
	negative skew;	B1dep (2)
(e)	mean = $(55-5)\times 0.9$	M1
	= 45	A1
	sd = 10×0.9	M1
	= 9	A1 (4) **Total 13**

NOTES	
(a)	B1 60 only
(b)	Award each B1 for correct answer only in this order.
(c)	M1 for use of correct formula, including square root. Correct answers with no working B1M1A1.
(d)	B1 any correct comparison of a pair of mean, median and mode using their values. B1 for 'negative skew' or allow (almost) symmetrical dependent upon correct reason.
(e)	M1 for (55 or 55.5 - 5)×0.9 A1 for the correct answer only. M1 for (10 or 10.3 or 10.5))×0.9 A1 for the correct answer only.

Q12.

Question Number	Scheme	Marks
(a)	Width = 0.5 (cm) e.g 4 [cm^{-2}] represents 8 babies or frequency densities are 8 and 34 Height = 17 (cm)	B1 M1 A1 (3)
(b)	$[Q_2 =]\ \{3\} + \frac{(25-9)}{(26-9)} \times 0.5$, or $\{3.5\} - \frac{(25-24)}{(41-24)} \times 0.5$ = awrt **3.47** (allow $\frac{59}{17}$)	M1, A1 (2)
(c)(i)	$\sum fx = 1\times1 + 2.5\times8 + 3.25\times17 + 3.75\times17 + 4.5\times7 = 171.5$, $\bar{x} = \frac{171.5}{50} = (3.43)$ (*)	B1cso
(ii)	$\sqrt{\frac{611.375}{50} - 3.43^2}$,= 0.680147… = awrt **0.680** (Accept 0.68)	M1, A1 (3)
(d)	$[P(W < 3) = P\left(Z < \frac{-0.43}{0.65}\right)] = P(Z < -0.6615..)$ $= 1 - 0.7454$ (tables) $= 0.2546$ awrt **0.254~0.255**	M1 M1 A1 (3)
(e)	(b) and (c)(i) mean ≠ med or skew or mean~ median or no skew **and** comment (d) = 0.254 or 0.255 compare data = 0.18 (or 12.7 compared with 9) 0.18 different from 0.25 so normal not good or 0.18 similar to 0.25 so normal is OK	B1 B1 dB1 (3)
(f)(i) (ii)	No change in mean (since weight is the same) s.d. will decrease (Extra value is at "centre" so data more concentrated) Both statements correct and correct reasons for each	B1 B1 dB1 (3) **[17 marks]**

	Notes	
(a)	M1 for clear representation of area with frequency or height × width = 8.5 A1 for 17 (cm) [Must be clear it is height not frequency] (Ans only must satisfy $h \times w$=8.5)	
(b)	M1 for $\frac{16}{17} \times 0.5$ or if using $n+1$ for $\frac{16.5}{17} \times 0.5$ May see $-\frac{1}{17} \times 0.5$ if working down A1 for awrt 3.47 (or $\frac{59}{17}$) [check from correct working] or (if using $(n+1)$ for 3.485 or awrt 3.49)	
(c)(i)	B1cso for Σfx (at least 3 correct & no incorrect products seen) and correct $\frac{\sum fx}{50}$ or $\frac{171.5}{50}$	
(ii)	M1 for a correct expression including square root. Must use 3.43 no ft A1 for awrt 0.680 (accept 0.68). Allow use of s = awrt 0.687 (Ans only 2/2)	
(d)	1st M1 for an attempt to standardise with 3, 3.43 and 0.65. Allow ± and also use of their sd 2nd M1 for $1-p$ where $0.74 < p < 0.75$ NB calculator gives 0.7458665… A1 for awrt 0.254 or 0.255	
(e)	1st B1 for a statement about mean/median **and** compatible comment about normal 2nd B1 for statement comparing their (d) with data (sight of 0.18 or 12.7 and 9 required) 3rd dB1 dep on 2nd B1 for conclusion about normal compatible with 2nd statement	
(f)(i) (ii)	1st B1 for no change in mean {send a correct argument for decrease to review} 2nd B1 for s.d. decreases 3rd dB1 dep on 1st and 2nd Bs for a correct reason for both mean and sd e.g. "new mean the same so within 1 s.d. of old mean"	

Q13.

Question Number	Scheme	Marks
(a)	4.5	B1 **(1)**
(b)	<table><tr><th>Time</th><th>Frequency density</th></tr><tr><td>2-4</td><td>5</td></tr><tr><td>5-6</td><td>4.5</td></tr><tr><td>7</td><td>6</td></tr><tr><td>8</td><td>24</td></tr><tr><td>9-10</td><td>7</td></tr><tr><td>11-15</td><td>2.4</td></tr></table> f.d = 24 is represented as 6cm, so f.d. = 7 is represented as **1.75(cm)**	M1 A1 A1 **(3)**
(c)	$\frac{1}{3}\times 15+9+\frac{1}{2}\times 6,=17$	M1, A1 **(2)**
(d)	$\text{Median} = 7.5+\frac{40-30}{24}\times 1 = 7.91666.....$ **awrt** 7.92 or 7.93(75) $Q_1 = 4.5+\frac{20-15}{9}\times 2 = 5.6111111...$ **awrt** 5.61 or 5.66(666…) $Q_3 = 8.5+\frac{60-54}{14}\times 2 = 9.357142.....$ **awrt** 9.36 or 9.46(4285….)	M1 A1 A1 A1 **(4)**
(e)	$Q_3 - Q_2 (= 1.4 \text{ or } 1.5) < Q_2 - Q_1 (= 2.3)$ or (Mean) < Median < Mode Therefore negative skew	B1ft dB1cao **(2)** **[Total 12]**

	Notes
(a)	B1 for 4.5 (o.e.) only. NB 1.5~4.5 is B0
(b)	M1 for evidence of f/w (at least 3 f.d. found). May be implied by a correct answer. A1 for identifying 9-10 as 2nd highest bar from correct working e.g. $24x = 6\times 7$ A1 for 1.75(cm). Correct answer only 3/3
(c)	M1 for a correct expression. May interpolate e.g. $\left[24+\frac{1}{2}\times 6-\frac{2}{3}\times 15\right]$ or (27 – 10) A1 for 17
(d)	M1 for one correct fraction in an expression for Q_1, Q_2 or Q_3 1st A1 for Q_2 **awrt** 7.92 (or 7.94 if (n +1) used – look for 40.5 instead of 40) 2nd A1 for Q_1 **awrt** 5.61 (or 5.67 if (n +1) used – look for 20.25 instead of 20) 3rd A1 for Q_3 **awrt** 9.36 (or 9.46 if (n +1) used – look for 60.75 instead of 60) NB watch out for working down e.g. $8.5-\frac{14}{24}\times 1$ for Q_2
(e)	1st B1ft for a correct comparison of their quartiles e.g. Q_2 closer to Q_3 or using at least two of Mean < Median < Mode (must state mean or mode if using this method). N.B. Mean = 7.71875, mode = 8 2nd B1cao dependent on 1st B1 being awarded for **negative skew** only (no ft)

	Notes	
(a)	B1 for 4.5 (o.e.) only. NB 1.5~4.5 is B0	
(b)	M1 for evidence of f/w (at least 3 f.d. found). May be implied by a correct answer. A1 for identifying 9-10 as 2nd highest bar from correct working e.g. $24x = 6\times 7$ A1 for 1.75(cm). Correct answer only 3/3	
(c)	M1 for a correct expression. May interpolate e.g. $\left[24+\frac{1}{2}\times 6-\frac{2}{3}\times 15\right]$ or $(27-10)$ A1 for 17	
(d)	M1 for one correct fraction in an expression for Q_1, Q_2 or Q_3 1st A1 for Q_2 **awrt** 7.92 (or 7.94 if $(n+1)$ used – look for 40.5 instead of 40) 2nd A1 for Q_1 **awrt** 5.61 (or 5.67 if $(n+1)$ used – look for 20.25 instead of 20) 3rd A1 for Q_3 **awrt** 9.36 (or 9.46 if $(n+1)$ used – look for 60.75 instead of 60) NB watch out for working down e.g. $8.5-\frac{14}{24}\times 1$ for Q_2	
(e)	1st B1ft for a correct comparison of their quartiles e.g. Q_2 closer to Q_3 or using at least two of Mean < Median < Mode (must state mean or mode if using this method). N.B. Mean = 7.71875, mode = 8 2nd B1cao dependent on 1st B1 being awarded for **negative skew** only (no ft)	

Q14.

Question Number	Scheme	Marks
(a)	Width = 4 (cm)	B1
	Area of 14 cm^2 represents frequency 28 and area of $4h$ represents 18	M1
	Or $\frac{4h}{18}=\frac{14}{28}$ (o.e.) h = **2.25** (cm)	A1
		(3)
(b)	$m=(240)+\frac{10}{22}\times 80$ (o.e.)	M1
	$=276.36...$ $(\frac{3040}{11})$ ((£)**276** $\leq m <$ (£)276.5)	A1
		(2)
(c)	$\sum fy = 31600$ leading to $\bar{y}=316$	M1A1
	$\sigma_y=\sqrt{\frac{12452800}{100}-(\bar{y})^2}$ = 157 .07... (awrt **157**) Allow s = 157.86...	M1A1
		(4)
(d)	Skewness = 0.764... (awrt **0.76** or **0.75**)	B1
	[If n+1 used in (b) and m = £278 accept awrt 0.73 or 0.72]	
	Positive skew	B1ft
		(2)
(e)	$z=\pm\frac{80}{150}$	M1
	P(240< X < 400) = **0.40 ~ 0.41**	A1
		(2)
(f)	(e) suggests a reasonable fit for this range BUT	B2/1/0
	(d) since skew it will not be a good fit overall	(2)
		15

	Notes
(a)	B1 for width (ignore units)
	M1 for clear method using area and frequency or their width × their height = 9
	e.g. seeing both fd of 0.7 and 0.225 (may see fd in the table) [Must use correct interval]
(b)	M1 for $\frac{10}{22}\times 80$ or $\frac{10.5}{22}\times 80$ (o.e.). Allow use of (n + 1) leading to £278 .18... or [278, 278.5)
	A1 Do not award if incorrect end-point seen but answer only is 2/2
(c)	1st M1 attempt at $\sum fy$ with at least 3 correct products or ans. that rounds to 30 000 (to 1 sf) &/100
	2nd M1 for correct expression including $\sqrt{}$. Follow through $\bar{y}$. Need $\sum fy^2$ correct but condone a minor transcription error e.g. 12458200.
(d)	1st B1 for awrt 0.76/0.75 for m = £276 or awrt 0.73/0.72 for m = £278
	2nd B1ft for a correct description of their skew based on their measure or if no measure given based on their values of mean and median. (correlation is B0)
(e)	M1 for an attempt to standardise using the 320 and 150 and either 240 or 400 (implied by 0.53)
	A1 for answer in range [0.40 , 0.41] (tables gives 0.4038, calculator 0.40619...) Ans only 2/2
(f)	For B2 we need 2 comments that make reference to each of part (e) and part (d)
	One comment should suggest it is not good since skew. The other it is since matches range in (e)
	1st B1 for one relevant comment
	2nd B1 for both comments NB Do not use B0B1

Q15.

Question	Scheme	Marks
(a)	One large square $= \frac{450}{"22.5"}$ or one small square $= \frac{450}{"562.5"}$ (o.e. e.g. $\frac{"562.5"}{450}$)	M1
	One large square = 20 cars or one small square = 0.8 cars or 1 car = 1.25 squares	A1
	No. > 35 mph is: $4.5\times"20"$ or $112.5\times"0.8"$ (or equivalent e.g. using fd)	dM1
	= **90** (cars)	A1 (4)
(b)	$[\overline{x}] = \frac{30\times12.5+240\times25+90\times32.5+30\times37.5+60\times42.5}{450}$ $\left[=\frac{12975}{450}\right]$	M1 M1
	$= 28.83...$ or $\frac{173}{6}$ **awrt 28.8**	A1 (3)
(c)	$[Q_2 =]$ $20+\frac{195}{240}\times10$ (o.e.) [Allow use of $(n+1)$ giving 195.5 instead of 195]	M1
	$= 28.125$ [Use of $(n+1)$ gives 28.145…] **awrt 28.1**	A1 (2)
(d)	$Q_2 < \overline{x}$ So positive skew | [Condone $Q_2 \approx \overline{x}$] [so (almost) symmetric]	B1ft dB1ft (2)
(e)	[If chose skew in (d)] **median (Q_2)** Since the data is skewed or median not affected by extreme values | [If chose symmetric in (d)] **mean ($\overline{x}$)** Since it uses all the data	B1 dB1 (2)
		[13]

	Notes
(a)	1st M1 for attempt to count squares (accept "22.5" in [22, 23] and "562.5" in [550, 575]) **and** use 450 to obtain a measure of scale. [If using fd must use 450 to obtain scale factor]
	1st A1 for a correct calc. for 20 or 0.8 or 1.25 etc
	[May be fd = 4 to 1 large sq. or 0.8 to 1 small sq. May be on the diagram.]
	2nd dM1 dep on 1st M1 for correctly counting squares for > 35 mph and forming suitable expr'
	2nd A1 for 90 with no incorrect working seen.
	e.g. $\frac{4.5}{22.5}\times450$ scores M1A1M1 and A1 when = 90 is seen. Answer only is 4/4
(b)	1st M1 for clear, sensible use of mid-points at least 3 of (12.5, 25, 32.5, 37.5, 42.5) seen
	2nd M1 for an expression for $\overline{x}$ (at least 3 correct terms on num' and a compatible denominator)
	Follow through their frequencies.
	You may see these fractions: $\frac{16218.75}{562.5}$ (small squares), $\frac{12975}{450}$ (frequencies), $\frac{648.75}{22.5}$ (large squares)
	A1 for awrt 28.8 (answer only is 3/3)
(c)	M1 for a full expression for median (using their frequencies). May see e.g. $25+\frac{75}{120}\times5$ etc
	Do nor accept boundaries of 19.5 or 20.5, these are M0A0
	A1 for awrt 28.1 (answer only is 2/2) [For use of $(n+1)$ accept 28.15 but not 28.2]
(d)	1st B1ft for a correct statement about their Q_2 and $\overline{x}$ [Condone $Q_2 \approx \overline{x}$ only if $\lvert Q_2 - \overline{x}\rvert < 1$]
	Do not accept an argument based on the shape of the graph alone.
	2nd dB1ft dependent on 1st B1 for a compatible description of skewness. F.t. their values
Quartiles	If $Q_1 = 23.4$ and $Q_3 = 33.7 \sim 33.8$ are seen allow comparison of quartiles for 1st B1 in (d)
(e)	1st B1 for a correct choice based on their skewness comment in (d). If no choice made in (d) only Q_2
	2nd dB1 for a suitable compatible comment

Q16.

Question Number	Scheme	Marks
(a)	$\underline{10.5}$	B1 (1)
(b)	$(Q_2 =)\ (15.5+)\ \frac{\frac{1}{2}\times 30-14}{8}\times 3$ or $\frac{\frac{1}{2}\times 31-14}{8}\times 3$	M1
	$= \underline{15.875 \text{ or } 16.0625}$	A1 (2)
(c)	$\bar{x} = \frac{477.5}{30} = \underline{15.9}$ (15.916&) [Accept $\frac{191}{12}$ or $15\frac{11}{12}$]	M1, A1
	$\sigma = \sqrt{\frac{8603.75}{30} - \bar{x}^2}$ $= \underline{5.78}$ (accept $s = 5.88$)	M1A1ft, A1 (5)
(d)	Since <u>mean and median are similar (or equal or very close)</u> a normal distribution may be suitable. [Allow mean or median close to <u>mode/modal class</u>]	B1 (1)
(e)	$Q_3 - Q_2 (=8) > (4.5 =) Q_2 - Q_1$	M1
	Therefore <u>positive skew</u>	A1 (2)
		(11 marks)

Notes

In parts (a) to (c) a correct answer with no working scores full marks for that value.

(a)
B1 for 10.5 which may be in the table

(b)
M1 for a correct ratio and times 3, ignore the lower boundary for this mark
A1 for awrt 15.9 (if n =30 used) or awrt 16.1 (if n+1 = 31 is used)

(c)
1st M1 for attempt at $\sum fx$ (this may be seen in the table as fx: 10, 73.5, 70, 136, 82, 106 [condone 1 slip] or awrt 500) and use of $\frac{\sum fx}{\sum f}$ or a correct expression for mean.
1st A1 for awrt 15.9
2nd M1 for an attempt at σ or σ^2, can ft their mean, condone mis-labelling $\sigma^2 = \sqrt{...}$ etc
Allow use of their $\sum fx^2$ (awrt 9000)
2nd A1ft for a correct expression including square root, ft their mean but not their $\sum fx^2$.
No label or correct label is OK but wrong label (e.g. $\sigma^2 = \sqrt{...}$) is A0
3rd A1 for awrt 5.78, allow s = awrt 5.88. **SC** Allow M1A1A0 for awrt 5.79 if $\bar{x}$ correct

(d)
B1 for a reason implying or stating symmetry. "Time is continuous" or "evenly distributed" is B0

(e)
M1 for a clear reason or comparison, values not essential but comparison implying they have been found is required.
A1 for stating "positive skew". Condone just "positive" but "positive correlation" is A0
Do not allow arguments based on mean and median since this part relates to a different set of data.

Q17.

Question Number	Scheme	Marks
(a)	Median = 32/2 = 16th term (16.5) $\frac{x-39.5}{49.5-39.5}=\frac{16-14}{25-14}$ or $x=39.5+\left(\frac{2}{11}\times 10\right)$ Median = 41.3 (use of $n+1$ gives 41.8) **(awrt 41.3)**	M1 A1 (2)
(b)	Mean$=\frac{1414}{32}=44.1875$ **(awrt 44.2)** Standard deviation $=\sqrt{\frac{69378}{32}-\left(\frac{1414}{32}\right)^2}$ $= 14.7$ (or $s = 14.9$)	B1 M1 A1 (3)
(c)	mean > median therefore positive skew	B1ft B1ft (2) [7]

Notes

(a)

M1 for an attempt to use interpolation to find the median. Condone use of 39 or 40 for 39.5
e.g. allow $39+\frac{2}{11}\times 10$ (o.e.) or $40+\frac{2}{11}\times 10$ (o.e.) to score M1A0 but must have the 10
A1 for awrt 41.3 (or awrt 41.8 if using $(n + 1)$)

(b)

B1 for awrt 44.2
M1 for a correct expression including square root. (Allow ft of their mean)
A1 for awrt 14.7 (If using s for awrt 14.9)
You may see $\sum t = 1339 \rightarrow \bar{t} = 41.8$ and $\sum t^2 = 62928 \rightarrow \sigma$ 14.7 or $s = 14.9$

Mid-points

this scores B0 for the mean but can score M1 for a correct st.dev expression and A1 for ans.
Correct answer only in (a) and (b) can score full marks but check (n +1) case in (a)

(c)

1st B1ft for a correct comparison of their mean and their median (may be in a formula)
Calculating median – mean as negative is OK for this B1 but must say +ve skew for 2nd B1
Only allow comparison to be ≈ 0 if $|\text{mean} - \text{median}| \leq 0.5$

2nd B1ft for a correct description of skewness based on their values of mean and median.
ft their values for mean and median not their previous calculation/comparison
Must be compatible with their previous comparison (if they have one)
"Positive skew" with no reason is B0B1 provided you can see their values that imply that.
Description should be "positive" or "negative" or "no" skew or "symmetric"
"Positive correlation" is B0

Quartiles

1st B1ft if Q_1 = awrt 32 and Q_3 = awrt 49 seen and a correct comparison made. ft Q_2

2nd B1ft if Q_1 = awrt 32 or Q_3 = awrt 49 seen and a correct description based on their quartiles and their comparison is made. (Should get "negative skew")

Q18.

Question Number	Scheme	Marks
(a)	23, 35.5 (may be in the table)	B1 B1 (2)
(b)	Width of 10 units is 4 cm so width of 5 units is **2 cm**	B1
	Height = 2.6×4 = **10.4 cm**	M1 A1 (3)
(c)	$\sum fx = 1316.5 \Rightarrow \bar{x} = \frac{1316.5}{56} =$ awrt **23.5**	M1 A1
	$\sum fx^2 = 37378.25$ can be implied	B1
	So $\sigma = \sqrt{\frac{37378.25}{56} - \bar{x}^2} =$ awrt **10.7** allow $s = 10.8$	M1 A1 (5)
(d)	$Q_2 = (20.5) + \frac{(28-21)}{11}\times 5 = 23.68\ldots$ awrt **23.7 or 23.9**	M1 A1 (2)
(e)	$Q_3 - Q_2 = 5.6,\ Q_2 - Q_1 = 7.9$ (or $\bar{x} < Q_2$)	M1
	[7.9 >5.6 so] **negative skew**	A1 (2)
		Total 14

(b) M1 for their width x their height=20.8.
Without labels assume width first, height second and award marks accordingly.

(c) 1st M1 for reasonable attempt at $\sum x$ and /56
2nd M1 for a method for σ or s, $\sqrt{}$ is required
Typical errors $\sum (fx)^2 = 354806.3$ M0, $\sum f^2x = 13922.5$ M0 and $(\sum fx)^2 = 1733172$ M0
Correct answers only, award full marks.

(d) Use of $\sum f(x-\bar{x})^2 =$ awrt 6428.75 for B1
lcb can be 20, 20.5 or 21, width can be 4 or 5 and the fraction part of the formula correct for M1 - Allow 28.5 in fraction that gives awrt 23.9 for M1A1

(e) M1for attempting a test for skewness using quartiles or mean and median.
Provided median greater than 22.55 and less than 29.3 award for M1 for $Q_3 - Q_2 < Q_2 - Q_1$ without values as a valid reason.
SC Accept mean close to median and no skew oe for M1A1

Q19.

Question Number	Scheme	Marks
(a)	70 – 80 group - width **0.5** (cm)	B1
	1.5 cm^2 is 10 customers <u>or</u> 3.75cm^2 is 25 customers <u>or</u> $0.5c = 3.75$ <u>or</u> $\frac{2.5}{\frac{1}{5}}$	M1
	70 – 80 group - height **7.5** (cm)	A1
		(3)
(b)	Median $= (70) + \frac{13.5}{25} \times 10$ allow $(n+1) = (70) + \frac{14}{25} \times 10$	M1
	$=$ **75.4** (or if using $(n+1)$ allow 75.6)	A1
		(2)
(c)	$\left[\text{Mean} = \frac{6460}{85}\right] =$ **76**	B1
	$\sigma = \sqrt{\frac{529400}{85} - 76^2}$	M1
	$= 21.2658\ldots\ldots$ $(s = 21.3920)$ **awrt 21.3**	A1
		(3)
(d)	Coeff` of skewness $= \frac{3(76-75.4)}{21.2658\ldots} = 0.08464\ldots$ **awrt 0.08** (awrt 0.06 for 75.6)	M1 A1
	There is (very slight) positive skew or the data is almost symmetrical (or both) <u>Any</u> mention of "correlation" is B0	B1ft
		(3)
		Total 11

	Notes
(a)	B1 for 0.5
	M1 for one of the given statements <u>or</u> any method where "their width" × "their height" = 3.75 Correct height scores M1A1 independent of width so B0M1A1 is possible.
(b)	M1 for a correct fraction: $+\frac{k}{25} \times 10$ where $k = 13.5$ or 14 for $(n+1)$ case.
	NB may work down so look out for (80) $-\frac{11.5}{25} \times 10$ etc **Beware:** $69.5 + \frac{13.5}{25} \times 11 = 75.44$ (but M0)
(c)	M1 for a correct expression with square root, ft their mean A1 for awrt 21.3 or, if clearly using s allow awrt 21.4. Must be evaluated...no surds.
(d)	M1 sub. their values into formula allow use of s but their σ or s must be > 0 A1 for awrt 0.08 but accept 0.085 No fraction B1ft for a correct comment compatible with their coefficient. Allow "symmetrical" for |coeff`| < 0.25 They may say it is "slightly skew" so omit "positive" but do not allow "negative" if coef` +ve Condone "strongly" positive skew.

Q20.

Question Number	Scheme	Marks
(a)	$P(D > 20) = P\left(Z > \frac{20-30}{8}\right)$	M1
	$= P(Z > -1.25)$	A1
	$= \underline{\mathbf{0.8944}}$ **awrt** $\underline{\mathbf{0.894}}$	A1 (3)
(b)	$P(D < Q_3) = 0.75$ so $\frac{Q_3 - 30}{8} = 0.67$	M1 B1
	$Q_3 =$ **awrt** $\underline{\mathbf{35.4}}$	A1 (3)
(c)	$35.4 - 30 = 5.4$ so $Q_1 = 30 - 5.4 =$ **awrt** $\underline{\mathbf{24.6}}$	B1ft (1)
(d)	$Q_3 - Q_1 = 10.8$ so $1.5(Q_3 - Q_1) = 16.2$ so $Q_1 - 16.2 = h$ or $Q_3 + 16.2 = k$	M1
	h=$\underline{\mathbf{8.4 \text{ to } 8.6}}$ and k= $\underline{\mathbf{51.4 \text{ to } 51.6}}$ both	A1 (2)
(e)	$2P(D > 51.6) = 2P(Z > 2.7)$	M1
	$= 2[1 - 0.9965] =$ awrt $\underline{\mathbf{0.007}}$	M1 A1 (3)
		Total 12

(a) M1 for an attempt to standardise 20 or 40 using 30 and 8.
1st A1 for $z = \pm 1.25$
2nd A1 for awrt 0.894

(b) M1 for $\frac{Q_3 - 30}{8}$ = to a z value
M0 for 0.7734 on RHS.
B1 for (z value) between 0.67~0.675 seen.
M1B0A1 for use of $z = 0.68$ in correct expression with awrt 35.4

(c) Follow through using their of quartile values.

(d) M1 for an attempt to calculate 1.5(IQR) and attempt to add or subtract using one of the formulae given in the question - follow through their quartiles

(e) 1st M1 for attempting 2P($D >$ their k) or (P($D >$their k)+ P($D <$ their h))
2nd M1 for standardising their h or k (may have missed the 2) so allow for standardising P($D > 51.6$) or P($D < 8.4$)
Require boths Ms to award A mark.

Q21.

Question Number	Scheme		Marks
Note : if a correct answer is given with no incorrect working award full marks unless the markscheme says otherwise.			
(a)	Mean = 1.41	B1: Cao Allow 141/100	B1
	Variance = $\frac{343}{100} - 1.41^2$	M1: using $\frac{\sum fx^2}{100} - (\text{their mean})^2$ or $\frac{100}{99}\left(\frac{\sum fx^2}{100} - (\text{their mean})^2\right)$ oe NB Allow the square root of this for the M mark. If no working shown for $\sum fx^2$ then you must see 343, 3.43 or a correct answer	M1
	= 1.4419 ($s^2 = 1.456$)	A1: awrt 1.44 or 1.46 for s^2	A1
			(3)
(b)	The mean is close to the variance	B1: Cao - allow alternative wording Allow mean equals variance	B1 (1)
(c) (i)	X~Po(1.5) $P(X = 2) = \frac{e^{-1.5}1.5^2}{2!}$	M1: writing or using $\frac{e^{-\lambda}\lambda^2}{2!}$ or $P(X \le 2) - P(X \le 1)$	M1
	= 0.2510	A1: awrt 0.251	A1
(ii)	$P(X \ge 1) = 1 - P(X = 0)$ $= 1 - e^{-1.5}$ 0r $1 - 0.2231$	M1: writing or using $1 - P(X = 0)$ oe	M1
	= 0.77686….	A1: awrt 0.777	A1 (4)
(d)	Y~Po(7.5)	B1: Writing Po(7.5)	B1
	$P(Y \ge 11) = 1 - P(Y \le 10)$	M1: writing $P(Y \ge 11)$ or $1 - P(Y \le 10)$ oe	M1
	$= 1 - 0.8622$ $= 0.1378$ *	A1: Seeing $1 - 0.8622$ leading to 0.1378 cso (both B1 and M1awarded)	A1cso (3)
(e)	A~ B(12, 0.1378)	M1: using $(p)^n(1-p)^{12-n}$ where p = 0.1378 or 0.138 condone missing nCr	M1
	$P(A = 3) = \binom{12}{3}(0.1378)^3(0.8622)^9$	M1: $\binom{12}{3}(p)^3(1-p)^9$, with $0 < p < 1$ Allow 220 or 12 C 3 instead of $\binom{12}{3}$	M1
	= 0.1516	A1: awrt 0.152	A1 (3)
			Total 14

Q22.

Question Number	Scheme	Notes	Marks
(a)	$\bar{x} = \frac{60}{15} = 4$	4 cao	B1
	$s^2 = \frac{1}{14}\left(1946 - 15 \times 4^2\right) = 121.857...$	M1 Use of complete, correct formula and attempt to substitute. A1 awrt 122 or $\frac{853}{7}$	M1,A1
			(3)
(b)(i)	$\bar{x} \pm 1.96 \times \frac{10}{\sqrt{15}} = 4 \pm 5.06$	Accept use of $\bar{x} \pm z \times \frac{10 \text{ or "their } s\text{"}}{\sqrt{15}}$, A1 all correct. Accept $\bar{x} = 0835$.	M1,A1
	$(-1.06, 9.06)$	Can be implied from correct interval below.	A1
	$(08\,29\,56, 08\,40\,04)$	Accept $(08\,29.94, 0840.06)$ or expressed using words or as an inequality. Accept answers to the nearest minute ie (0830,0840).	A1
(ii)	Paul samples times of **buses randomly** or **independently** of each other	Context required.	B1
			(5)
(c)	0 / 0831 / 8.31(am) is 'contained in' the confidence interval	Award if comment about their interval is correct. Only accept 'above the lower limit of' etc if the statement taken as a whole clearly means 'contained in'.	M1
	Paul's belief is not supported / 0831 arrival time is reasonable	Must contain some context	A1cao
			(2)
			Total 10